

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 11:56:36 ; Search time 10354 Seconds  
(without alignments)  
11513.488 Million cell updates/sec

Title: US-09-843-007A-1  
Perfect score: 2914  
Sequence: 1 GAGTTTTCGCTCCGAACC.....CGTTTCAGACGGCATTTCG 2914

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenBank

- 1: gb\_ba.\*
- 2: gb\_hvg.\*
- 3: gb\_in.\*
- 4: gb\_on.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_on.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2641	90.6	2883	6	A47369	A47369 Sequence 1
2	2641	90.6	2883	6	A47785	A47785 Sequence 1
3	2093	71.8	2115	1	NPO011781	AJ011781 Neisseria
4	1921.4	65.9	1939	6	AX107013	AX107013 Sequence
5	1921.4	65.9	6878	6	AX107015	AX107015 Sequence
6	1891.4	64.9	1910	6	AX107014	AX107014 Sequence
7	1891.4	64.9	6851	6	AX107016	AX107016 Sequence
8	1853.4	63.6	1967	1	AY099334	AY099334 Neisseria
9	1847	63.4	1967	1	AY099335	AY099335 Neisseria
10	655	22.5	795	6	AX300255	AX300255 Sequence
11	655	22.5	329861	1	NMA522491	AL162756 Neisseria
12	639	21.9	765	6	AX565901	AX565901 Sequence
13	639	21.9	13001	1	AE002494	AE002494 Neisseria
14	639	21.9	349980	6	AX044032	AX044032 Sequence
15	380.2	13.0	10719	1	AE001946	AE001946 Deinococc
16	314	10.8	11605	1	AE012454	AE012454 Xanthomon
17	303.8	10.4	10261	1	AE011997	AE011997 Xanthomon
18	239	8.2	10822	1	AE005791	AE005791 Caulobact
19	170.6	5.9	10644	1	AE005392	AE005392 Escherich
20	170.6	5.9	10838	1	AE015192	AE015192 Shigella
21	170.6	5.9	257071	1	AP002558	AP002558 Escherich
22	170.6	5.9	290582	1	AE016984	AE016984 Shigella
23	169	5.8	759	1	ECARODG	X59503 E.coli aroD
24	169	5.8	759	1	SDU82269	U82269 Shigella dy
25	169	5.8	10685	1	AE000264	AE000264 Escherich
26	169	5.8	19521	1	D90811	D90811 E.coli geno
27	164.2	5.6	300475	1	AE016761	AE016761 Escherich
28	163.8	5.6	1798	1	ECAROD	X04306 E. coli aro
29	163	5.6	941	1	STAROD3D	X54546 S. typhi ar
30	161.4	5.5	233050	1	AL627271	AL627271 Salmonell
31	161.4	5.5	300523	1	AE016838	AE016838 Salmonell
32	157.8	5.4	759	1	AF184963	AF184963 Salmonell
33	157.8	5.4	20279	1	AE008759	AE008759 Salmonell
34	156.4	5.4	420	1	NGU39803	U39803 Neisseria g
35	153.2	5.3	249050	1	AL596165	AL596165 Listeria
36	153.2	5.3	349980	6	AX417038	AX417038 Sequence
37	153.2	5.3	349980	6	AX417041	AX417041 Sequence
38	150	5.1	280050	1	AL591975	AL591975 Listeria
39	150	5.1	349980	6	AX641665	AX641665 Sequence
40	147.4	5.1	1340	6	AX416059	AX416059 Sequence
41	147.4	5.1	189050	1	AL646077	AL646077 Ralstonia
42	142.6	4.9	300029	1	AE016952	AE016952 Enterococ
43	141	4.8	2684	1	ENEBESA	L23802 Enterococcu
44	133	4.6	756	6	AX414335	AX414335 Sequence
45	129.2	4.4	28206	1	BACDIA	L09228 Bacillus su

ALIGNMENTS

RESULT 1  
A47369  
LOCUS A47369 2883 bp DNA linear PAT 07-MAR-1997  
DEFINITION Sequence 1 from Patent WO9531553.  
ACCESSION A47369  
VERSION A47369.1 GI:2301369  
KEYWORDS Neisseria polysacchara  
SOURCE Neisseria polysacchara  
ORGANISM Neisseria polysacchara  
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
Neisseriaceae; Neisseria.  
REFERENCE 1 (bases 1 to 2883)  
Kossmann, J., Buettcher, V. and Welsh, T.  
AUTHORS DNA SEQUENCES CODING FOR ENZYMES CAPABLE OF FACILITATING THE  
SYNTHESIS OF LINEAR alpha -1,4 GLUCANS IN PLANTS, FUNGI AND

Pred. No. is the number of results predicted by chance to have a

```

JOURNAL      MICROORGANISMS
PATENT: WO 9531553-A 1 23-NOV-1995;
INST GENBIOLOGISCHE FORSCHUNG (DE)
COMMENT      Other publication DE 4447388 960627
              Other publication AU 2614195 951205.
              Location/Qualifiers
FEATURES
    source    1..2883
              /organism="Neisseria polysaccharia"
              /mol_type="genomic DNA"
              /db_xref="taxon:489"
              /clone="PNB2"
              /clone_lib="GENOMIC LIBRARY IN PLBUESCRIPTII SK"
              /note="unnamed protein product"
              /codon_start=1
              /transl_table=11
              /protein_id="CAA03058.1"
              /db_xref="GI:2301370"
              /translation="MLTPQQVGLILQYKTRILDIYTPQAGIEKSEDMRQFRRM
DTHFKLMNEDSVVGNBHALPMLLELAQWQSYQRNSLSLKDIDDIARENNPWIL
SNKQVGVVYDLFAGDLKGLDKIYFQELGLYILHMLPLFKPSPKSDGGYAVSY
RDVNPALGTIGDLRIRVIAALHSHFRRRFYLPQHLQRTMAQRACAGPLDNFYIIF
PRRMPDQYDRLRLRIPDPQHPGFSOLEDRWVTNSPMDILNPNVFAOWRA
KCSPLTRALITSCVMWMLPLFGNKHQAAKTCAAHALIPAFNAVMEIAAPAVFKEEA
IVHPQVVGITIGDCQOIGTNELQWALLNLTATREVNLLHQALTYRNLPEHTAWN
YVSHDDIGWTFAISDAALVIGSYDHRQFLNRFVNRFPDGTFAFGVFPQINPSTGDC
RVSQTAALVLAQDDPHAVDRIKLLYSIALSTGLGLPLYLGLDEVLTDDDDWCDAAI
RATTAAGVPSLQSPRATERSVDRSQIYQGLRHMIAVROSFRPFGDGLVNTNTN
NKHIIGYATMRFMSVTSANIRKPLRIPCKPSPSRRTTSSVAKLSA"
BASE COUNT  561 a  852 c  758 g  612 t
ORIGIN

Query Match      90.6%; Score 2641; DB 6; Length 2883;
Best Local Similarity 97.8%; Pred. No. 0;
Matches 2852; Conservative 0; Mismatches 30; Indels 33; Gaps 16;

QY  1  GAGTTTGGTTCCTCCGAAACGAAAGTGATGCTTGAGCCGAAACACCTGTCCGCGCAAGCGG 60
DB  1  GAGTTTGGTTCCTCCGAAACGAAAGTGATGCTTGAGCCGAAACACCTGTCCGCGCAAGCGG 60
QY  61  CTGACCGCCCTTTTGCCCAATCGACATCGTAAACAAATCGTTGGTGCAAGCTCTTTTC 120
DB  61  -----ACCGCTTTTGCCCAATCGACATCGTAAACAAATCGTTGGTGCAAGCTCTTTTC 114
QY  121  GCTTCAGCGGTGCGAAGACGAAAGTCAGACGCTTTCGCGCTTTCGCGCATCACCGCA 180
DB  115  GCTTCAGCGGTGCGAAGACGAAAGTCAGACGCTTTCGCGCTTTCGCGCATCACCGCA 173
QY  181  ATTTTCAGATGTCGCGCGCGAGTCCTCCATCTGTTTCAGACGCGATACGATTTCTTCT 240
DB  174  ATTTTCAGATGTCGCGCGCGAGTCCTCCATCTGTTTCAGACGCGATACGATTTCTTCT 233
QY  241  TCGCGCGCGGTGGGTGAACCTCATGATTCGACAGCGAGCGCGGATCGGTTTTCGA 300
DB  234  TCGCGCGCGGTGGGTGAACCTCATGATTCGACAGCGAGCGCGGATCGGTTTTCGA 290
QY  301  GCATTCGCCACCGCGCGCGGAGCGCGGTTTCGCCGGAAGAAAGCTCGATATCGATAATG 360
DB  291  GCATTCGCCACCGCGCGG-----AGCGGTTTCGCCGGAAGAAAGCTCGATATCGATAATG 344
QY  361  TCGGGAAGCGCGGCTTTCATTCAGCGAGTCGACGAGTTCAAAATAATTCGTCGGAACAC 420
DB  345  TCGGGAAGCGCGGCTTTCATTCAGCGAGTCGACGAGTTCAAAATAATTCGTCGGAACAC 404
QY  421  GGGAAACGAGCGCGCTTCGCGCATGCGCTCTGAAAGTAAACGACGAGCGCGCTTTCGCGCAGC 480
DB  405  GGGAAACGAGCGCGCTTCGCGCATGCGCTCTGAAAGTAAACGACGAGCGCGCTTTCGCGCAGC 464
QY  481  GCCTTCGCGGACGCTCTGCGGTGCGGCAATACCTTCGCGGATGCTCCCGCGGATTCGAAA 540
DB  465  GCCTTCGCGGACGCTCTGCGGTGCGGCAATACCTTCGCGGATGCTCCCGCGGATTCGAAA 524
QY  541  AAATCGCGCGGAACCTCGACGATATCGAAGGCGGAGTTTGTGATTTGGTCAAGTACGCGG 600

```

```

DB  525  AAATCGCGCGGAACCTCGACGATATCGAAGGCGAGGTTTTTGATTTGTCACAGTACGCGG 584
QY  601  GAAAGTACGCGCGCATTCGCGGCGACAAAGCGGACCGGATTTTGGTGGCTCGCTTCG 660
DB  585  GAAAGTACGCGCGCATTCGCGGCGACAAAGCGGACCGGATTTTGGTGGCTCGCTTCG 644
QY  661  ATAACGGTGTTTTGAACGGTCA-GGCTGGTGTGCATGCGCGTGTGTGCGGTGAAAGGAA 719
DB  645  ATAACGGTGTTTTGAACGGTCAAGGGCTGGTGTGCATGCGCGTGTGTGCGGTGAAAGGAA 704
QY  720  CGGTAAAGACCAATTTATGAAAGGACAGGCGAATTTTCAGACGCGCATTTCTGTGCGG 779
DB  705  CGGTAAAGACCAATTTATGAAAGGACAGGCGAATTTTCAGACGCGCATTTCTGTGCGG 764
QY  780  CCGGCTTGATTAATCAAGCAGCATTCGCGATATCGGAATCGACATTTGGCAAGCCCT 839
DB  765  CCGGCTTGATTAATCAAGCAGCATTCGCGATATCGGAATCGACATTTGGCAAG-CCT 823
QY  840  GTCTTTTCTAGTCAGTCGCGAGTTCTTTCAGTATGATTTGACGACACGCGCCCTACACGCA 899
DB  824  GTCTTTTCTAGTCAGTCGCGAGTTCTTTCAGTATGATTTGACGACACGCGCCCTACACGCA 883
QY  900  TTTCAGGATACGCGCGGACAGACCGCGTCCGAAACCTTCAGAAATCGGAGCAGGCATCATG 959
DB  884  TTTCAGGATACGCGCGGACAGACCG--GTCCGAAATTCAGAAATCGGAGCAGGCATCATG 941
QY  960  TTGACCCCCACGACAGAAAGTCGTTGATTTTACGATACCTCAAAAACAGCATCTTGAC 1019
DB  942  TTGACCCCCACGACAGAAAGTCGTTGATTTTACGATACCTCAAAAACAGCATCTTGAC 1001
QY  1020  ATCTACGCGCGCAACAGCGCGCGCATTCGAAATTCGAAAGACTGCGCGGCGAGTTTCG 1079
DB  1002  ATCTACGCGCGCAACAGCGCGCGCATTCGAAATTCGAAAGACTGCGCGGCGAGTTTCG 1061
QY  1080  CGCGCATGATAGCATTTTCCCAAACTGATGAAAGAACTGCGACAGCGGTGTAACGCAAC 1139
DB  1062  CGCGCATGATAGCATTTTCCCAAACTGATGAAAGAACTGCGACAGCGGTGTAACGCAAC 1121
QY  1140  AACGAAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1199
DB  1122  AACGAAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1181
QY  1200  CAAACGCAACTCATCTTAAAGATATCGATATCGCGCGCGGCGGAAACAAACCCGATGGAAT 1259
DB  1182  CAAACGCAACTCATCTTAAAGATATCGATATCGCGCGCGGCGGAAACAAACCCGATGGAAT 1241
QY  1260  TTGTCCAAACAAAGTTCGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1319
DB  1242  TTGTCCAAACAAAGTTCGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1301
QY  1320  GCTTGAAGATATAAATTCCTTATTTTCAAGAGCTTGGTTTGAATTTATCTGCACTGATG 1379
DB  1302  GCTTGAAGATATAAATTCCTTATTTTCAAGAGCTTGGTTTGAATTTATCTGCACTGATG 1361
QY  1380  CGCTGTTTAAATGCCCTGAAAGGCAAAAGCGCGGCTATGCGGTGCGAGCTACCGC 1439
DB  1362  CGCTGTTTAAATGCCCTGAAAGGCAAAAGCGCGGCTATGCGGTGCGAGCTACCGC 1421
QY  1440  GATGTCATTCGCGCACTGCGGCAATAGCGCACTTTCGCGCAAGTCAATGTCGCTGCGTGCAC 1499
DB  1422  GATGTCATTCGCGCACTGCGGCAATAGCGCACTTTCGCGCAAGTCAATGTCGCTGCGTGCAC 1481
QY  1500  GAAGCGCGCATTTTCGCGGTGCTGATTTTATCTTCAACCAACACCTTCCAAAGCAACGAA 1559
DB  1482  GAATC--GCATTTTCGCGGTGCTGATTTTATCTTCAACCAACACCTTCCAAAGCAACGAA 1539
QY  1560  TGGGCGCAACGCTGCGCGCGCGGCGGCGGCGGCTTTTCGCAATTTCTACTATATTTTCCC 1619
DB  1540  T--GGGCGCAACGCTG--GGCGCGGCGGCGGCTTTTCGCAATTTCTACTATATTTTCCC 1595
QY  1620  GACCGCGGATGCGCGCAATACGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1679

```

1596 GACCGCGGATGCCGACCAATACGACCGCACCTCTGGCGAAATCTTCCCGACGACAC 1655  
1680 CCGGGGGCTTCTCGCACTGGAAGACGACGACCTGGTGTGAGACACCTTCAATTCCTTC 1739  
1656 CCGGGGGCTTCTCGCACTGGAAGACGACGACCTGGTGTGAGACACCTTCAATTCCTTC 1715  
1740 CAATGGGACTTGAATTTACAGCAACCGTGGGTATTCGCGCAATCGCGGGGCAATGCTG 1799  
1716 CAATGGGACTTGAATTTACAGCAACCGTGGGTATTCGCGCAATCGCGGGGCAATGCTG 1774  
1800 TTCTTGGCAACTTGGGCGCTTGAATTCCTGCTGATGATGCGGTTCCTTTATTTGGAAA 1859  
1775 TTCTTGGCAACTTGGGCGCTTGAATTCCTGCTGATGATGCGGTTCCTTTATTTGGAAA 1834  
1860 CAATGGGACAGCTGCGAAACCTTCGCGGACGCGACGCTCTCATCGCGCTTCAAT 1919  
1835 CAATGGGACAGCTGCGAAACCTTCGCGGACGCGCTCTCATCGCGCTTCAAT 1892  
1920 GCGGTATCGGTATTCGCGGCGCGCGGTGTTCTTCAATCGGAAGCCATGTCACGCC 1979  
1893 GCGGTATCGGTATTCGCGGCGCGCGGTGTTCTTCAATCGGAAGCCATGTCACGCC 1952  
1980 GACCAAGTCTCAATACATCGGCGAGGAGCAATGCGCAATCGGTACAAACCCCTGCA 2039  
1953 GACCAAGTCTCAATACATCGGCGAGGAGCAATGCGCAATCGGTACAAACCCCTGCA 2012  
2040 ATGGCATTTCTTGGAAACACCTTCGCGGCGCGAGTCAACCTGCTCATCAGCGCTG 2099  
2013 ATGGCATTTCTTGGAAACACCTTCGCGGCGCGAGTCAACCTGCTCATCAGCGCTG 2072  
2100 ACTACCGGCAACCTTCGCGGAGCATACCGCTGGGTCACTAGCTCGGAGGACAGAC 2159  
2073 ACTACCGGCAACCTTCGCGGAGCATACCGCTGGGTCACTAGCTCGGAGGACAGAC 2132  
2160 GACATCGGCTGACGTTTCGCGATGAAGACGCGGATATCTGGGATAGCGCTACGAC 2219  
2133 GACATCGGCTGACGTTTCGCGATGAAGACGCGGATATCTGGGATAGCGCTACGAC 2192  
2220 CACCGGCAATTCCTCAACCGCTTCTTGGTCAACCGGTTTCGAGCGGAGCTTCTGCTG 2279  
2193 CACCGGCAATTCCTCAACCGCTTCTTGGTCAACCGGTTTCGAGCGGAGCTTCTGCTG 2252  
2280 GTACGGTTCATACCAACCGGAGCAGCGGAGTCCGCTGTCAGTGTACGCGCGGCA 2339  
2253 GTACGGTTCATACCAACCGGAGCAGCGGAGTCCGCTGTCAGTGTACGCGCGGCA 2312  
2340 TTGGTTCGGCTTGGCGAAGACGATCCCGGCTGACCGGATCAAACTCTTGTACAGC 2399  
2313 TTGGTTCGGCTTGGCGAAGACGATCCCGGCTGACCGGATCAAACTCTTGTACAGC 2372  
2400 ATTGCTTTGAGTACCGGCGCTTGGCGGCTGATTTACCTAGGCGAGGAGTGGTACGCTC 2459  
2373 ATTGCTTTGAGTACCGGCGCTTGGCGGCTGATTTACCTAGGCGAGGAGTGGTACGCTC 2432  
2460 AATGACGAGCTGTTGCGAAGACGATCAATAGAGCGAGCAGCGGCTTGGCGGACCGT 2519  
2433 AATGACGAGCTGTTGCGGAGCAGGAGCAGCGGCTTGGCGGACCGT 2490  
2520 CCGGCTCAACGAGCGCTTGTACGCGCAACGCAACGATCGGTGACGCGCGGCA 2579  
2491 CCGGCTCAACGAGCGCTTGTACGCGCAACGCAACGATCGGTGACGCGCGGCA 2549  
2580 ATCTACGAGCTTGGCGGATATGATTCGCGGCAAGCAATCGCGGCTTGTACGCGC 2639  
2550 ATCTACGAGCTTGGCGGATATGATTCGCGGCAAGCAATCGCGGCTTGTACGCGC 2609  
2640 GGCAGGCTGGTACATTTCAACCAACCAACGATCATCGGTACATCGCGCAACAT 2699  
2610 GGCAGGCTGGTACATTTCAACCAACCAACGATCATCGGTACATCGCGCAACAT 2668  
2700 GGCCTTTGGCATTCGGTAACTTACGAGATATCGCAACCGGTTACCGGATACCGCTG 2759  
2669 GGCCTTTGGCATTCGGTAACTTACGAGATATCGCAACCGGTTACCGGATACCGCTG 2728

QY 2760 CAAGCCATGCGCTTCAAGGCGCAGCAGCTCATCGCTGCTGCAAACTGTGAGCTGAATCAG 2819  
DB 2729 CAAGCCATGCGCTTCAAGGCGCAGCAGCTCATCGCTGCTGCAAACTGTGAGCTGAATCAG 2788  
QY 2820 GATTGACGCTTCAGCCCTATCAGCTCATGTGCTGCGAAATCGCTGACGACGCTTCCC 2879  
DB 2789 GATTGACGCTTCAGCCCTATCAGCTCATGTGCTGCGAAATCGCTGACGACGCTTCCC 2848  
QY 2880 AAATGCGCTTCAACGCTTTCAGACGCGCATTTGCG 2914  
DB 2849 AAATGCGCTTCAACGCTTTCAGACGCGCATTTGCG 2883

## RESULT 2

A47785  
LOCUS A47785 2883 bp DNA linear PAT 07-MAR-1997  
DEFINITION Sequence 1 from Patent WO9533838.  
ACCESSION A47785  
VERSION A47785.1 GI:2301671  
KEYWORDS  
SOURCE  
ORGANISM

Neisseria polysaccharea  
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
Neisseriaceae; Neisseria.

1 (bases 1 to 2883)

REFERENCE  
AUTHORS Kossman, J., Buettcher, V. and Welsh, T.  
TITLE MICROORGANISMS PERMITTING THE INTRACELLULAR POLYHYDROXY ALKANOATE  
SYNTHESIS WITH SIMULTANEOUS EXTRACELLULAR POLYSACCHARIDE SYNTHESIS  
AND PROCESSES FOR PRODUCING THE SAME

JOURNAL Patent: WO 9533838-A 1 14-DEC-1995;  
INSTR GENBIOLOGISCHE FORSCHUNG (DE)  
COMMENT Other publication AU 2787895 960104  
Other publication DE 4420223 950504.

## FEATURES

## source

1. 2883  
/organism="Neisseria polysaccharea"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:489"  
/clone="PNB2"  
/clone\_lib="GENOMIC LIBRARY IN PBLUESCRIPTII SK"  
939. .2783  
/note="unnamed protein product"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA03083.1"  
/db\_xref="GI:2301672"  
/translation="MLPTQGVGLIQLYKTRILDIYTPQRIAGIEKSEDMRQFERRM  
DTHPKLNLSDVYGNNEALLPMLLEALQWQSYQNSLSKIDIDTARENPPMIL  
SNKQGVYVDLPAAGDLKGLKIPYQELGLTYLHMLPLFKPEKSDGGYAVSTY  
RDVNPALGTIDLEIVIAALHESHRERRFYLOPHLQRTMAQRCAGDPLDFNYIF  
PDRWPQDYDTLREIIPDHPGGSQLEDGRWYWTNFSQWMDLANSNPVAFPESEA  
KCCSLPTWALTSCVWMLPLFGNKWQAATCAHALIRAFNAVRIAPAVPESEA  
IVHPQVQYIGQDECGIYNPLQMLNLTATREVNLHQAITYIRHNLPEHTANVN  
YVRSHDIDGTFADDAAYLIGSYDHRQFLNRFVNRPDGTFANGVFQYNPFGDC  
RVSGTAAALVGLAQDDPHAVDRILKLYSIALSTGLPLIYLGDVETGLDNDWQAAI  
RAATVAGPSALQSPVRATERSVDRSQRIYQGLRHMEAVRQSNRPFQDGLVTFNTN  
NKHIIIGVIAIWRFNHVSITSANIRKPLRCPKPCPSRRTTSSVAKLSA"

## CDS

BASE COUNT 661 a 852 c 758 g 612 t  
ORIGIN  
Query Match 90.6%; Score 2641; DB 6; Length 2883;  
Best Local Similarity 97.8%; Pred. No. 0;  
Matches 2852; Conservative 0; Mismatches 30; Indels 33; Gaps 16;

QY 1 GAGTTTTCGTTCCCGAACCGAACGTTGATGCTGTGAGCCGACACCTGTCCGCAAGCGG 60  
DB 1 GAGTTTTCGTTCCCGAACCGAACGTTGATGCTGTGAGCCGACACCTGTCCGCAAGCGG 60  
QY 61 CTGACCGCGCCCTTTGCGCCCATCGACATCGTAACTCGGTTGCTGGCGCAAGCTCTTTC 120  
DB 61 -----ACCGCTTTGCGCCCATCGACATCGTAACTCGGTTGCTGGCGCAAGCTCTTTC 114





```
Db 2253 GTACCGTTCACATACACCAAGCAGCGGACTGCCGTGTCAGTGTGATACAGCCCGGCGA 2312
Qy 2340 TTGGTGGCTTGGCCGACAGCATCCCAAGCGGTGACCGCATCAAACTCTGTACAGC 2399
Db 2313 TTGGTGGCTTGGCCGACAGCATCCCAAGCGGTGACCGCATCAAACTCTGTACAGC 2372
Qy 2400 ATTGCTTTGAGTACCGCGGCTGCGGCTGATTTACCTAGGCGAGAGTGGTACGCTC 2459
Db 2373 ATTGCTTTGAGTACCGCGGCTGCGGCTGATTTACCTAGGCGAGAGTGGTACGCTC 2432
Qy 2460 AATGACGACGCTGTCGACAGCAGCAATAAGAGCAGCAGCGGTTGGCGCACCGT 2519
Db 2433 AATGACGACGCTGTCGACAGCAGCAATAAGAGCAGCAGCGGTTGGCGCACCGT 2490
Qy 2520 CCGCGCTCAACGAGCGCTGTCGCGGCAACGCAAGCATCGTCCGACGCGCGGCA 2579
Db 2491 CCGCGCTCAACGAGCGCTGTCGCGGCAACGCAAGCATCGTCCGACGCGCGGCA 2549
Qy 2580 ATCTATCAGGGCTTGGCCATATGATTCGCGTCCGCAAGCAATCCGCGCTTCGACGCG 2639
Db 2550 ATCTATCAGGGCTTGGCCATATGATTCGCGTCCGCAAGCAATCCGCGCTTCGACGCG 2609
Qy 2640 GCGAGCTGTTACATTTAAACCAACACACAGCAGCATCGCTACATCCGCAACAT 2699
Db 2610 GCGAGCTGTTACATTTAAACCAACACACAGCAGCATCGCTACATCCGCAACAT 2668
Qy 2700 GCGCTTTGGCATTCGTTAACTTCAGCGAATATCCGCAACCGTTACCGCGCATACCCCTG 2759
Db 2669 GCGCTTTGGCATTCGTTAACTTCAGCGAATATCCGCAACCGTTACCGCGCATACCCCTG 2728
Qy 2760 CAAGCATGCTTCAAGCGCAGACCTCATCGGTGCGCAAACTCGTACGCTGAATCAG 2819
Db 2729 CAAGCATGCTTCAAGCGCAGACCTCATCGGTGCGCAAACTCGTACGCTGAATCAG 2788
Qy 2820 GATTTGAGCTTCAGCGCTCATCAGCTCATGCGCTCGAAATCGCTGACGACGCTTCC 2879
Db 2789 GATTTGAGCTTCAGCGCTCATCAGCTCATGCGCTCGAAATCGCTGACGACGCTTCC 2848
Qy 2880 AAATCCGCTGTAACCGTTTCAGCGCATTTTGG 2914
Db 2849 AAATCCGCTGTAACCGTTTCAGCGCATTTTGG 2883

RESULT 3
NPO011781 2115 bp DNA linear BCT 28-JAN-1999
LOCUS Neisseria polysaccharea amylosucrase gene.
DEFINITION Neisseria polysaccharea amylosucrase gene.
ACCESSION AJ011781
VERSION AJ011781.1 GI:4049491
KEYWORDS amylosucrase; amylosucrase gene.
SOURCE Neisseria polysaccharea
ORGANISM Neisseria polysaccharea
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE 1
AUTHORS De Montalk, G.P., Renaud-Simeon, M., Willemot, R.M., Planchot, V. and
Monsan, P.
TITLE Sequence analysis of the gene encoding amylosucrase from Neisseria
polysaccharea and characterization of the recombinant enzyme
JOURNAL J. Bacteriol. 181 (2), 375-381 (1999)
MEDLINE 99102197
PUBMED 9882648
REFERENCE 2 (bases 1 to 2115)
AUTHORS Potocki de Montalk, G.
TITLE Direct submission
JOURNAL Submitted (02-OCT-1998) Potocki de Montalk G., U.M.R. C.N.R.S.
5504, L.A. I.N.R.A. D.G.B.A., I.N.S.A., Centre de biogenierie
Gilbert Durand, Complexe scientifique de Rangueil, Toulouse cedex,
31077, FRANCE
FEATURES
source
1..2115
/organism="Neisseria polysaccharea"
/mol_type="genomic DNA"
```

```
/strain="ATCC 43768"
/db_xref="taxon:489"
136..2046
/gene="amylosucrase"
<136..2046
/gene="amylosucrase"
/EC_number="2.4.1.4"
/codon_start=1
/transl_table=11
/product="amylosucrase"
/protein_id="CAA09772.1"
/db_xref="GI:4107260"
/transl_table="MLPTQOQVILQVLRILDIQTPQORAGIERSKEDWRQFSRRM
DTHPKLNEILDSVYGNELPLMLLQAWOSVDSORNSLXLDIDIDARNRNPWIL
SNKQVGVYVDLPAQDLKGLDKIPLYOELGLTYLHMLPKPKGKSGDGYAVSY
RDVNPALGTTIGDLREVLIAALHEAGISAVNDIFFNHTSNEHEWAORCAAGDPLFNFYI
IPDRMPDQYDRLREIFPDQHRGFSQVEDQWVWTFNSFQMDLNSNPWFVFRAM
AGEKFLANGLRDLRMDAVAFIWKQGTSCENLPQAHALIRAFNVRIRAPVFFPK
SEATVHPDDQVQYIGDEQCTGYNPLQWALLMNTLATREVNLLHQAITYRENLEPHTA
WNTVYRHDDIGTFAEDDAVYLGISGDRHQLNRFVNRFDGSGFARGVFLNDWPSQ
GDCRVSTAAALVGLAODDPAVDRIKLKYSIALSTGCLPLIYLGDEVNPRFDGSLVT
DSKSDSRWHRPRYNEALYAQRNDPSTAGQIYQGLRHMIAVQSNRPRFDGSLVT
FNTNNKILGIRNALLAFENFSEYFQVTAHTLQAMPFKAHDLIGKTVSLNQDLT
IQPVQVWMLLEIA"
BASE COUNT 493 a 661 c 529 g 432 t
ORIGIN
Query Match 71.8%; Score 2093; DB 1; Length 2115;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2093; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 822 AGACTTGGCAACAAGCCCTGCTTTCTTAGTCAGTCGCGAGTCTTGCAATGATGTCAC 881
Db 1 AGACTTGGCAACAAGCCCTGCTTTCTTAGTCAGTCGCGAGTCTTGCAATGATGTCAC 60
Qy 882 GACGCGCTTACAGCGCATTTGCGAGGATACGCGCGCAGACGCGCGTGGGAACTTCAGA 941
Db 61 GACGCGCTTACAGCGCATTTGCGAGGATACGCGCGCAGACGCGCGTGGGAACTTCAGA 120
Qy 942 ATCGGAGCAGGCATCATGTTGACCCCGCAGCAGCAGTCTGTTGATTTACAGTACCTC 1001
Db 121 ATCGGAGCAGGCATCATGTTGACCCCGCAGCAGCAGTCTGTTGATTTACAGTACCTC 180
Qy 1002 AAAACAGCATCTTGGACATCTACGCGCGAAGCAGCGCGCGCATCGGAAATCCGAA 1061
Db 181 AAAACAGCATCTTGGACATCTACGCGCGAAGCAGCGCGCGCATCGGAAATCCGAA 240
Qy 1062 GACTGGCGGCGAGTTTTCGCGCGCATGATGATACGCAATTTCCCAAACTGATGAAAGACTC 1121
Db 241 GACTGGCGGCGAGTTTTCGCGCGCATGATGATACGCAATTTCCCAAACTGATGAAAGACTC 300
Qy 1122 GACAGGTGTACGCGCAACAGAGAGCCCTGCTGCTATGCTGGAATGCTGCTGCGCGAG 1181
Db 301 GACAGGTGTACGCGCAACAGAGAGCCCTGCTGCTATGCTGGAATGCTGCTGCGCGAG 360
Qy 1182 GCATGCGAAAGCTATTCGCAAGCAATCTCATCTCTTAAAGATATCGATATCGCGCGAA 1241
Db 361 GCATGCGAAAGCTATTCGCAAGCAATCTCATCTCTTAAAGATATCGATATCGCGCGAA 420
Qy 1242 AACCAACCCGATTTGCTTGTCCAAACAAAGTCCGCGCGCTGCTGCTGCTGCTGCTGCTG 1301
Db 421 AACCAACCCGATTTGCTTGTCCAAACAAAGTCCGCGCGCTGCTGCTGCTGCTGCTGCTG 480
Qy 1302 TTTGCGCGGCTTGAAGGCTTGAAGATTAATTCCTTATTTTCAAGAGCTTGGTTTG 1361
Db 481 TTTGCGCGGCTTGAAGGCTTGAAGATTAATTCCTTATTTTCAAGAGCTTGGTTTG 540
Qy 1362 ACTTATCTGACCTGATGCGCGCTGTTTAAATGCCCTGAAGCAAAAGCAGCGCGCTAT 1421
Db 541 ACTTATCTGACCTGATGCGCGCTGTTTAAATGCCCTGAAGCAAAAGCAGCGCGCTAT 600
Qy 1422 GCGGTGAGGCTACCGGATGTCATTCGCGCATCGGCAATAGGAGCTTGGCGAA 1481
```

Db 601 GCGGTGAGGAGTACCGGATGTAATCGGCACTGGGCAATAGGCGACTTGGCGAA 660  
Qy 1482 GTCAATGTCGCTGCAAGAGCGGCAATTCGCGCTGTGTAATTTATTTCAACCAAC 1541  
Db 661 GTCAATGTCGCTGCAAGAGCGGCAATTCGCGCTGTGTAATTTATTTCAACCAAC 720  
Qy 1542 ACCTCAAGCAACGAAATGGGCGCAAGCTGCGCGCGGCGGACCGCTTTTCGACAA 1601  
Db 721 ACCTCAAGCAACGAAATGGGCGCAAGCTGCGCGCGGCGGACCGCTTTTCGACAA 780  
Qy 1602 TTCTACTATATTTTCCCGCAACCGCGGATGCGCGCAATACGACCGCAACCTTCGCGAA 1661  
Db 781 TTCTACTATATTTTCCCGCAACCGCGGATGCGCGCAATACGACCGCAACCTTCGCGAA 840  
Qy 1662 ATCTTCCCGCAACGACCGCGGCTTCTCGCACTGCGCAAGCGCTGGGCTGG 1721  
Db 841 ATCTTCCCGCAACGACCGCGGCTTCTCGCACTGCGCAAGCGCTGGGCTGG 900  
Qy 1722 AGACCTTCAATTCCTTCCAAATGGGACTTGAATACAGCAACCGCTGGGTATTCGCGCA 1781  
Db 901 AGACCTTCAATTCCTTCCAAATGGGACTTGAATACAGCAACCGCTGGGTATTCGCGCA 960  
Qy 1782 ATGGCGGCGAATGCTTCTTCCAAATGGGACTTGAATACAGCAACCGCTGGGTATTCGCGCA 1841  
Db 961 ATGGCGGCGAATGCTTCTTCCAAATGGGACTTGAATACAGCAACCGCTGGGTATTCGCGCA 1020  
Qy 1842 GTTCCCTTTATTTGGAAACAAATGGGCAAGCTGCGAAACCTGCGGAGCGCAACGCG 1901  
Db 1021 GTTCCCTTTATTTGGAAACAAATGGGCAAGCTGCGAAACCTGCGGAGCGCAACGCG 1080  
Qy 1902 CTATCGCGCGTTCAAATGGGCTTATGCGTATTCGCGCGCGCGCGCTTTCGCAATCG 1961  
Db 1081 CTATCGCGCGTTCAAATGGGCTTATGCGTATTCGCGCGCGCGCGCTTTCGCAATCG 1140  
Qy 1962 GAAGCCATGCTTCCAAATGGGCAAGCTGCGAAACCTGCGGAGCGCAACGCG 2021  
Db 1141 GAAGCCATGCTTCCAAATGGGCAAGCTGCGAAACCTGCGGAGCGCAACGCG 1200  
Qy 2022 GGTACAAACCTTCCAAATGGGCTTATGCGTATTCGCGCGCGCGCTTTCGCAATCG 2081  
Db 1201 GGTACAAACCTTCCAAATGGGCTTATGCGTATTCGCGCGCGCGCTTTCGCAATCG 1260  
Qy 2082 CTCTCCATGAGCGCTGAGCAACGCGCAACCTGCGGAGCGCAACGCG 2141  
Db 1261 CTCTCCATGAGCGCTGAGCAACGCGCAACCTGCGGAGCGCAACGCG 1320  
Qy 2142 TAGCTCGCGCGCAACGCGCAACGCGCAACCTGCGGAGCGCAACGCG 2201  
Db 1321 TAGCTCGCGCGCAACGCGCAACGCGCAACCTGCGGAGCGCAACGCG 1380  
Qy 2202 GGCAATAGCGCTGAGCAACGCGCAACCTGCGGAGCGCAACGCG 2261  
Db 1381 GGCAATAGCGCTGAGCAACGCGCAACCTGCGGAGCGCAACGCG 1440  
Qy 2262 GGCAATAGCGCTGAGCAACGCGCAACCTGCGGAGCGCAACGCG 2321  
Db 1441 GGCAATAGCGCTGAGCAACGCGCAACCTGCGGAGCGCAACGCG 1500  
Qy 2322 AGTGGTACAGCGCGGCAATGGTTCGCTGCGCAAGCAACGCGCTTCGCG 2381  
Db 1501 AGTGGTACAGCGCGGCAATGGTTCGCTGCGCAAGCAACGCGCTTCGCG 1560  
Qy 2382 ATCAAACTTTGTACAGCAATTTGCTGAGTACCGCGCTTCTGCGCTGATTTACTAGG 2441  
Db 1561 ATCAAACTTTGTACAGCAATTTGCTGAGTACCGCGCTTCTGCGCTGATTTACTAGG 1620  
Qy 2442 GACCAAGTGGGTACGCTCAATGACGAGCTGCTGCGCAAGCAACGCAATAGAGCGAG 2501  
Db 1621 GACCAAGTGGGTACGCTCAATGACGAGCTGCTGCGCAAGCAACGCAATAGAGCGAG 1680  
Qy 2502 AGCGGTTGGGCGCAACGCGCTTACCAAGCGCTTGTACGCGCAACGCGCTTCG 2561  
Db 1681 AGCGGTTGGGCGCAACGCGCTTACCAAGCGCTTGTACGCGCAACGCGCTTCG 1740

Qy 2562 TCGACCGCGAGCGGCAATCTATCAGGCTTGGCGCAATATGATTCGCGCTCGCAAGC 2621  
Db 1741 TCGACCGCGAGCGGCAATCTATCAGGCTTGGCGCAATATGATTCGCGCTCGCAAGC 1800  
Qy 2622 AATCGCGCTTTCGAGCGGCGAGGCTGTTTACATTTCAACCAACCAAGCACATCATC 2681  
Db 1801 AATCGCGCTTTCGAGCGGCGAGGCTGTTTACATTTCAACCAACCAAGCACATCATC 1860  
Qy 2682 GGCTACATCGCAACCAATGCGCTTTTGGCATTCGGTACCTTCAGCGCAATATCCGCAAC 2741  
Db 1861 GGCTACATCGCAACCAATGCGCTTTTGGCATTCGGTACCTTCAGCGCAATATCCGCAAC 1920  
Qy 2742 GTTACCGGCGATACCTTCGAGCGCAATGCGCTTTCAGCGCGCAACCTTCATCGTGGCA 2801  
Db 1921 GTTACCGGCGATACCTTCGAGCGCAATGCGCTTTCAGCGCGCAACCTTCATCGTGGCA 1980  
Qy 2802 ACTGTACGCTGAATCAGGATTTGAGCTTTCAGCGCTTTCAGCGCGCAACCTTCATCGTGGCA 2861  
Db 1981 ACTGTACGCTGAATCAGGATTTGAGCTTTCAGCGCTTTCAGCGCGCAACCTTCATCGTGGCA 2040  
Qy 2862 GCTTCAAGCGCACTTCCCAATGCGCTTTCAGCGCGCAACCTTTCAGCGCGCAATTTGG 2914  
Db 2041 GCTTCAAGCGCACTTCCCAATGCGCTTTCAGCGCGCAACCTTTCAGCGCGCAATTTGG 2093

RESULT 4  
AX107013  
LOCUS AX107013 1939 bp DNA linear PAT 30-APR-2001  
DEFINITION Sequence 4 from Patent WO0125449.  
ACCESSION AX107013  
VERSION AX107013.1 GI:13922547  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM artificial sequences.

REFERENCE  
1 Bengs, H., Polakowski, T., Held, A. and Gallert, K. C.  
Novel immobilizable amylosucrase, the use thereof and method for  
producing poly(1,4-alpha-glucan)  
Patent: WO 0125449-A 4 12-APR-2001;  
Axiva GmbH (DE)

FEATURES  
Location/Qualifiers  
1..1939  
/organism="synthetic construct"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32630"  
/note="PCR-Amplikon, kodierend fur amylosucrase (AmSu)"  
BASE COUNT 453 a 605 c 482 g 395 t  
ORIGIN

Query Match 65.9%; Score 1921.4; DB 6; Length 1939;  
Best Local Similarity 99.9%; Pred. No. 0; Mismatches 1; Indels 0; Gaps 0;  
Matches 1922; Conservative 0;

Qy 955 TCATGTTGACCGCGCGAGCAAGCTCGGTTTGATTTTACAGTACCTCAAAACACGATCT 1014  
Db 10 TCATGTTGACCGCGCGAGCAAGCTCGGTTTGATTTTACAGTACCTCAAAACACGATCT 69  
Qy 1015 TGGACATCTACGCGCGCGAGCAAGCTCGGCGCGGATCGAAATTCGAGACTGCGCGAGT 1074  
Db 70 TGGACATCTACGCGCGCGAGCAAGCTCGGCGCGGATCGAAATTCGAGACTGCGCGAGT 129  
Qy 1075 TTTTCGCGCGCGATGGAATGCGATTTTCCCAAACTGATGAACCACTTCGACAGCGTACG 1134  
Db 130 TTTTCGCGCGCGATGGAATGCGATTTTCCCAAACTGATGAACCACTTCGACAGCGTACG 189  
Qy 1135 GCACCAACGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1194  
Db 190 GCACCAACGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 249  
Qy 1195 ATTCCCAACGCACTCATCTTAAAGATATCGATATCGCGCGCGCAAAACACCGCGATT 1254

Db	250	ATTCCCAACGCAATCTATCTCTTAAAGATATCGATATATCGCGCGCGAAGAAACAACCCCGGATT	309
Qy	1255	GGATTTTGTCCAAACAAACAAAGTCGGCGCGGTGTCTACGTTTGAATTTGTTTGCOCGCGATT	1314
Db	310	GGATTTTGTCCAAACAAACAAAGTCGGCGCGGTGTCTACGTTTGAATTTGTTTGCOCGCGATT	369
Qy	1315	TGAAGGCGTTGAAGAGATAAAATTCCTTATTTTCAAGAGCTTGGTTTGCATTTCTGCACCC	1374
Db	370	TGAAGGCGTTGAAGAGATAAAATTCCTTATTTTCAAGAGCTTGGTTTGCATTTCTGCACCC	429
Qy	1375	TGATGCCGCTGTGTTTAAATTCGCTGAAGGCAAAAGCGAGCGCGGCTATGTCGGTTCAGCAGCT	1434
Db	430	TGATGCCGCTGTGTTTAAATTCGCTGAAGGCAAAAGCGAGCGCGGCTATGTCGGTTCAGCAGCT	489
Qy	1435	ACCOCGATGTCATCTCCGCACTGTGGCAAAATAGCGCACTTTCGGCGAAGTCAATTCCTCGCC	1494
Db	490	ACCOCGATGTCATCTCCGCACTGTGGCAAAATAGCGCACTTTCGGCGAAGTCAATTCCTCGCC	549
Qy	1495	TGCACGAAGCCGGCATTTTCGCGCGTGTGCAATTTTATCTTCAACCAACACCTCCACGCAAC	1554
Db	550	TGCACGAAGCCGGCATTTTCGCGCGTGTGCAATTTTATCTTCAACCAACACCTCCACGCAAC	609
Qy	1555	ACGAATGGGCGCAACGCTTGCGCCCGCGCGCACCGGCTTTTCGACAAATTTCTACTATATTT	1614
Db	610	ACGAATGGGCGCAACGCTTGCGCCCGCGCGCACCGGCTTTTCGACAAATTTCTACTATATTT	669
Qy	1615	TCCCGGACCGCGATGCGCGCAACCAATACGACCGCACCGCTGCGCGAATCTTCCCGCAAC	1674
Db	670	TCCCGGACCGCGATGCGCGCAACCAATACGACCGCACCGCTGCGCGAATCTTCCCGCAAC	729
Qy	1675	AGCACCCGCGCGGCTTTCGCACTTGGAAGACGGACGCTGGGTGTGACGACCTTCAATTT	1734
Db	730	AGCACCCGCGCGGCTTTCGCACTTGGAAGACGGACGCTGGGTGTGACGACCTTCAATTT	789
Qy	1735	CTTTCGAAATGGGACTTGAATTAAGCAACCGGTGGGTATTTCCGCGCAATTCGCGGGGAGAA	1794
Db	790	CTTTCGAAATGGGACTTGAATTAAGCAACCGGTGGGTATTTCCGCGCAATTCGCGGGGAGAA	849
Qy	1795	TGCTGTTTCTTTCGCAACTTGGCGCTTGACATCTCTCGCTATGGATTCGGTTTGCCTTTATTT	1854
Db	850	TGCTGTTTCTTTCGCAACTTGGCGCTTGACATCTCTCGCTATGGATTCGGTTTGCCTTTATTT	909
Qy	1855	GGAAACAAATGGGGACAAAGCTGCGAAACCTGCGGAGGGCGACCGCTCATTCGCGCGCT	1914
Db	910	GGAAACAAATGGGGACAAAGCTGCGAAACCTGCGGAGGGCGACCGCTCATTCGCGCGCT	969
Qy	1915	TCAATGCCGTTATCGCTATTTCGCGCGCCGCGCGTGTCTTCAAAATCGAAGCCATCTGTC	1974
Db	970	TCAATGCCGTTATCGCTATTTCGCGCGCCGCGCGTGTCTTCAAAATCGAAGCCATCTGTC	1029
Qy	1975	ACCCGACCAAGTCGTCCAAATACATCGGGCAGGACGAATGCGAAATCGGTTACAAACCCC	2034
Db	1030	ACCCGACCAAGTCGTCCAAATACATCGGGCAGGACGAATGCGAAATCGGTTACAAACCCC	1089
Qy	2035	TGCAAAATGGCATGTTTGTGGAAACACCTTGCACCGCGGAAGTCAACCTGCTCCATCAGG	2094
Db	1090	TGCAAAATGGCATGTTTGTGGAAACACCTTGCACCGCGGAAGTCAACCTGCTCCATCAGG	1149
Qy	2095	CGCTGACCTACCGCCACAACCTGCCGAGCATACCGCTGGGTCAACTAGCTTCGAGACC	2154
Db	1150	CGCTGACCTACCGCCACAACCTGCCGAGCATACCGCTGGGTCAACTAGCTTCGAGACC	1209
Qy	2155	ACGACGACATCGGCTGGACGTTTTCGGATGAAGCGGCTATATCTGGGATTAAGCGCT	2214
Db	1210	ACGACGACATCGGCTGGACGTTTTCGGATGAAGCGGCTATATCTGGGATTAAGCGCT	1269
Qy	2215	ACGACCAACCGCAATTCCTCAACCGCTTCTTGTCAACCGTTTCGACGGCAGCTTCGCTC	2274
Db	1270	ACGACCAACCGCAATTCCTCAACCGCTTCTTGTCAACCGTTTCGACGGCAGCTTCGCTC	1329
Qy	2275	GTGCGCTACCGTTCCAAATACAAACCAAGCACAGCGGACTGCGGTGTCACTGCTACAGCCG	2334
Db	1330	GTGCGCTACCGTTCCAAATACAAACCAAGCACAGCGGACTGCGGTGTCACTGCTACAGCCG	1389

QY	2335	CGGCATTGGTGGCTGGCGCAAGAGCATGCCCAACGCCCGTTGACCCGCATCAAACTCTTGT	2339		
Db	1390	CGGCATTGGTGGCTGGCGCAAGAGCATGCCCAACGCCCGTTGACCCGCATCAAACTCTTGT	1449		
QY	2395	ACAGCATGCTTTGAGTACCGCGCGGCTCTGCCGCTGATTTTACCTTAGCGGACGAAAGTGGGTA	2454		
Db	1450	ACAGCATGCTTTGAGTACCGCGCGGCTCTGCCGCTGATTTTACCTTAGCGGACGAAAGTGGGTA	1509		
QY	2455	CGCTCAATGACGACGACTGGTTCGCAAGACAGCAATTAAGAGCGACGACAGCCGTTGGCGCG	2514		
Db	1510	CGCTCAATGACGACGACTGGTTCGCAAGAGCAATTAAGAGCGACGACAGCCGTTGGCGCG	1569		
QY	2515	ACGTCGCGCTCAACCGAAGCCCTGTACGCGCAACGCAACGATCCGTTCGACCGCAGCGG	2574		
Db	1570	ACGTCGCGCTCAACCGAAGCCCTGTACGCGCAACGCAACGATCCGTTCGACCGCAGCGG	1629		
QY	2575	GGCAAACTATACGGGCTTGGGCATATGATTTGCCGCTCGCCCAAGAGCAATCCGCGCTTCG	2634		
Db	1630	GGCAAACTATACGGGCTTGGGCATATGATTTGCCGCTCGCCCAAGAGCAATCCGCGCTTCG	1689		
QY	2635	ACGGCGGACGCTGGTTTACATTCACCAACCAACAAAGCACATCATCGGCTACATCCGCA	2694		
Db	1690	ACGGCGGACGCTGGTTTACATTCACCAACCAACAAAGCACATCATCGGCTACATCCGCA	1749		
QY	2695	ACAAATGCGTTTTGGCAATTCGGTAATCTCAGCGAATATCGCAAAACGTTACCGCGCAT	2754		
Db	1750	ACAAATGCGTTTTGGCAATTCGGTAATCTCAGCGAATATCGCAAAACGTTACCGCGCAT	1809		
QY	2755	CCCTGCAAGCCATGCGCTTCAAGCGCACGACCTCATCGTGGCAAACTGTGAGCCCTGA	2814		
Db	1810	CCCTGCAAGCCATGCGCTTCAAGCGCACGACCTCATCGTGGCAAACTGTGAGCCCTGA	1869		
QY	2815	ATCAGGATTTGACGCTTCAGCCCTATCAGGTCAATGGCTCGAAATCGGCTGACGCAACG	2874		
Db	1870	ATCAGGATTTGACGCTTCAGCCCTATCAGGTCAATGGCTCGAAATCGGCTGACGCAACG	1929		
QY	2875	TTC 2877			
Db	1930	CTC 1932			
RESULT 5					
LOCUS	AX107015	6878 bp	DNA linear PAT 30-APR-2001		
DEFINITION	Sequence 6 from Patent WO0125449.				
ACCESSION	AX107015				
VERSION	AX107015.1	GI:13922549			
KEYWORDS	synthetic construct				
SOURCE	synthetic construct				
ORGANISM	artificial sequences.				
REFERENCE	1				
AUTHORS	Bengs, H., Polakowski, T., Held, A. and Gallert, K.C.				
TITLE	Novel immobilizable amylosucrase, the use thereof and method for				
	producing poly(1,4-alpha-glucan)				
JOURNAL	Patent: WO 0125449-A 6 12-APR-2001;				
	Axiva GmbH (DE)				
FEATURES	Location/Qualifiers				
source	1...6878				
	/organism="synthetic construct"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:32630"				
	/note="Beschreibung der kunstlichen				
	Sequenz: Expressionsvektor pGEX-4T-1-AmSu"				
BASE COUNT	1673 a	1800 c	1766 g 1639 t		
ORIGIN					
Query Match	55.9%;	Score 1921.4;	DB 6; Length 6878;		
Best Local Similarity	99.3%;	Pred. No. 0;			
Matches 1922;	Conservative	0; Mismatches	1; Indels 0; Gaps 0;		
QY	955	TCATGTTGACCCCGACGAGCTGGGTTTGATTTTACAGTACCTCAAAACAGCGCATCT	1014		

943 TCATGTTGACCCCGACGAGCAAGTGGTTGATTTTACAGTACCTCAAAACAGCATCT 1002  
1015 TGGACATCTACACGCGCCGGAAGCGCGCGCATCGAAAAATTCGAAAGCTTGGCGGAGT 1074  
1003 TGGACATCTACACGCGCCGGAAGCGCGCGCATCGAAAAATTCGAAAGCTTGGCGGAGT 1062  
1075 TTTCGCGCGCATGGATAGCATTTTCCCAAACTGATGAAAGAACTTCGACACGCTGTACG 1134  
1063 TTTCGCGCGCATGGATAGCATTTTCCCAAACTGATGAAAGAACTTCGACACGCTGTACG 1122  
1135 GCAACAAAGCGCTTGTGCTTATGCTGGAAGATGCTGCTGGCGAGGATGGCAAGCT 1194  
1123 GCAACAAAGCGCTTGTGCTTATGCTGGAAGATGCTGCTGGCGAGGATGGCAAGCT 1182  
1195 ATTCCCAACGCACTCATCTTAAGAGATATGATATCGCGCGGAAACAAACCCGAT 1254  
1183 ATTCCCAACGCACTCATCTTAAGAGATATGATATCGCGCGGAAACAAACCCGAT 1242  
1255 GGAATTTGTCGAACAAAGCTGGCGGCTGTGCTACCTTGAATTTGTTGCGCGGAT 1314  
1243 GGAATTTGTCGAACAAAGCTGGCGGCTGTGCTACCTTGAATTTGTTGCGCGGAT 1302  
1315 TGAAGGCTTGAAGATTAATTCCTATTTTCAGAGCTTGGTTGATCTTATCTGCAAC 1374  
1303 TGAAGGCTTGAAGATTAATTCCTATTTTCAGAGCTTGGTTGATCTTATCTGCAAC 1362  
1375 TGAATGCGCTGTTAAATGCGCTTGAAGCAAAAGCGAGCGGCTATGCGGTGAGAGCT 1434  
1363 TGAATGCGCTGTTAAATGCGCTTGAAGCAAAAGCGAGCGGCTATGCGGTGAGAGCT 1422  
1435 ACCGCGATGTCGAATCGCGGCTTGGCGCAATAGCGCACTTGGCGAGTCAATGCTGCGC 1494  
1423 ACCGCGATGTCGAATCGCGGCTTGGCGCAATAGCGCACTTGGCGAGTCAATGCTGCGC 1482  
1495 TGCACGAAGCGCGCATTTCCGCGCTTGGTCAATTTATCTTCAACCACTTCCAAAGAAC 1554  
1483 TGCACGAAGCGCGCATTTCCGCGCTTGGTCAATTTATCTTCAACCACTTCCAAAGAAC 1542  
1555 ACGAATGGGCGCAAGCTGCGCGCGCGCGGACCGCGCTTTTGCACAAATTTCTACTATTT 1614  
1543 ACGAATGGGCGCAAGCTGCGCGCGCGCGGACCGCGCTTTTGCACAAATTTCTACTATTT 1602  
1615 TCCCGGACCGCGGATGCGCGGACCAATACGACGCGACCTTGGCGGAAATCTTCCCGGAC 1674  
1603 TCCCGGACCGCGGATGCGCGGACCAATACGACGCGACCTTGGCGGAAATCTTCCCGGAC 1662  
1675 AGCACCGCGCGGCTTCTCGCAACTGGAAGACGCACTGGGTGTGGAAGACCTTCAAT 1734  
1663 AGCACCGCGCGGCTTCTCGCAACTGGAAGACGCACTGGGTGTGGAAGACCTTCAAT 1722  
1735 CCTTCAATGGGACTTGAATTAAGCAACCGGTGGGTATTCGCGCAATGGGCGGGA 1794  
1723 CCTTCAATGGGACTTGAATTAAGCAACCGGTGGGTATTCGCGCAATGGGCGGGA 1782  
1795 TGGTGTTCCTTGGCAACTTGGGCGGTTGACATCTCTGCTATGATGAGTGGCTTTATTT 1854  
1783 TGGTGTTCCTTGGCAACTTGGGCGGTTGACATCTCTGCTATGATGAGTGGCTTTATTT 1842  
1855 GGAACAAATGGGGAAGCTTGGCAACTTGGGCGGCGGCGGCGCTCATCGCGGCT 1914  
1843 GGAACAAATGGGGAAGCTTGGCAACTTGGGCGGCGGCGGCGCTCATCGCGGCT 1902  
1915 TCAATGCGGTATGCGTATTTGCGCGCGCGCGCTGTTCTTCAAAATCGGAGCCATCGTC 1974  
1903 TCAATGCGGTATGCGTATTTGCGCGCGCGCGCTGTTCTTCAAAATCGGAGCCATCGTC 1962  
1975 ACCCGGACCAAGTGTGTCATATCATGCGGCGGAGGAGGATGCGAAATCGGTTACAGCCGC 2034  
1963 ACCCGGACCAAGTGTGTCATATCATGCGGCGGAGGAGGATGCGAAATCGGTTACAGCCGC 2022  
2035 TGCATATGGCAATGTTGTGGAACACCTTGCACGCGCGAGTCAACCTGCTCATCAGG 2094

2023 TGCAATGSGCATTTGTTGGAAACACCTTGCACGCGGAGTCAAOCTGCTCCATCAGG 2082  
2095 CGCTGACCTACCGCCACCAACTGCTCCGAGCATACGCTGGGTCAAATCTAGCTCCGAGCC 2154  
2083 CGCTGACCTACCGCCACCAACTGCTCCGAGCATACGCTGGGTCAAATCTAGCTCCGAGCC 2142  
2155 ACGACGATCATCGGTGGAAGCTTTCGCAATGAAGACGCGGCAATCTTGGGCAATAGCGGCT 2214  
2143 ACGACGATCATCGGTGGAAGCTTTCGCAATGAAGACGCGGCAATCTTGGGCAATAGCGGCT 2202  
2215 ACGACGATCATCGGTGGAAGCTTTCGCAATGAAGACGCGGCAATCTTGGGCAATAGCGGCT 2274  
2203 ACGACGATCATCGGTGGAAGCTTTCGCAATGAAGACGCGGCAATCTTGGGCAATAGCGGCT 2262  
2275 GTGGCGTACCGTTCCAAATPACAAACCAAGCACAGCGACTGCGGTCTCATGCTGACAGCG 2334  
2263 GTGGCGTACCGTTCCAAATPACAAACCAAGCACAGCGACTGCGGTCTCATGCTGACAGCG 2322  
2335 CGGCATTTGCTCGGCTTGGCGCAAGCATCCCGCGGCTTGGCGCAATCAAACTCTTGT 2394  
2323 CGGCATTTGCTCGGCTTGGCGCAAGCATCCCGCGGCTTGGCGCAATCAAACTCTTGT 2382  
2395 ACAGCATTTGCTTGAATACCGGCGCTCTGCGGCTGATTTACTAGGCGACGAGTGGTA 2454  
2383 ACAGCATTTGCTTGAATACCGGCGCTCTGCGGCTGATTTACTAGGCGACGAGTGGTA 2442  
2455 CGCTCAATGACGACGCTGCTGCGCAAGCATGAGCAATGAGCGGACGACAGCGGTTGGCGC 2514  
2443 CGCTCAATGACGACGCTGCTGCGCAAGCATGAGCAATGAGCGGACGACAGCGGTTGGCGC 2502  
2515 ACCGTCGCGGCTTACAAACGAAAGCCCTGTACGCGCAACGCAACGATCCGTCGACCGAGCG 2574  
2503 ACCGTCGCGGCTTACAAACGAAAGCCCTGTACGCGCAACGCAACGATCCGTCGACCGAGCG 2562  
2575 GCGAATCTATCAGGCTTGGCGCAATGATTTGCGCTCGCGCAAGCAATCGCGCTTCG 2634  
2563 GCGAATCTATCAGGCTTGGCGCAATGATTTGCGCTCGCGCAAGCAATCGCGCTTCG 2622  
2635 ACGCGGCGAGGCTGTTACATTTCAACACCAACGCAACGATCATCGCTACATCCGCA 2694  
2623 ACGCGGCGAGGCTGTTACATTTCAACACCAACGCAACGATCATCGCTACATCCGCA 2682  
2695 ACAATGCGCTTTGGCAATTCGGTAACTTCAGGCAATATCCGCAACCGTTACCGGCGATA 2754  
2683 ACAATGCGCTTTGGCAATTCGGTAACTTCAGGCAATATCCGCAACCGTTACCGGCGATA 2742  
2755 CCTCGAAGCATGCTTCAAGCGCGACCTCATCGGTGGCAAACTGTCACCTGA 2814  
2743 CCTCGAAGCATGCTTCAAGCGCGACCTCATCGGTGGCAAACTGTCACCTGA 2802  
2815 ATCAGGATTTGACGCTTTCAGGCTTATCAGGCTATGCTGCTGAAATCGCTTCAACGCG 2874  
2803 ATCAGGATTTGACGCTTTCAGGCTTATCAGGCTATGCTGCTGAAATCGCTTCAACGCG 2862  
2875 TTC 2877  
2863 CTC 2865

RESULT 6  
AX107014

LOCUS AX107014  
DEFINITION Sequence 5 from Patent WO0125449.  
ACCESSION AX107014  
VERSION AX107014.1 GI:13922548  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM artificial sequences.  
REFERENCE 1  
AUTHORS Bengs, H., Polakowski, T., Held, A. and Gallert, K.C.  
TITLE Novel immobilizable amylase, the use thereof and method for producing poly(1,4-alpha-glucan)

AX107014 1910 bp DNA linear PAT 30-APR-2001  
Sequence 5 from Patent WO0125449.

JOURNAL Patent: WO 0125449-A 5 12-APR-2001;  
Axiva GmbH (DE)  
FEATURES Location/Qualifiers  
source 1. 1910  
/organism="synthetic construct"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32630"  
/note="PCR-Amplikon, kodierend fur verkurzte amylosucrase  
(AmSu5)"  
BASE COUNT 447 a 595 c 474 g 387 t 7 others  
ORIGIN  
Query Match 64.9%; Score 1891.4; DB 6; Length 1910;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1892; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 985 TGAATTTACAGTACTCAAAACACGACATCTTGGACATCTACACGCCGGAACAGCGCGCG 1044  
Db 11 TGAATTTACAGTACTCAAAACACGACATCTTGGACATCTACACGCCGGAACAGCGCGCG 70  
Qy 1045 GCATCGAAAAATCCGAAGACTGGCGGCGAGTTTTCGGCGCGCATGGATACGCATTTCCCCA 1104  
Db 71 GCATCGAAAAATCCGAAGACTGGCGGCGAGTTTTCGGCGCGCATGGATACGCATTTCCCCA 130  
Qy 1105 AACTGATGAACGAACTCGACAGCGTGTACGGCAACGACGAGCCCTGCTGCTATGCTGG 1164  
Db 131 AACTGATGAACGAACTCGACAGCGTGTACGGCAACGACGAGCCCTGCTGCTATGCTGG 190  
Qy 1165 AAATGCTGCTGGCGAGGCATGGCAAGCTATTCCCAACGCAACTCATCTCTTAAAGATA 1224  
Db 191 AAATGCTGCTGGCGAGGCATGGCAAGCTATTCCCAACGCAACTCATCTCTTAAAGATA 250  
Qy 1225 TCGATATCGCGCGGAAACGACCGGATTTGGATTTTGTCCAAACAAAGTGGCGCGCG 1284  
Db 251 TCGATATCGCGCGGAAACGACCGGATTTGGATTTTGTCCAAACAAAGTGGCGCGCG 310  
Qy 1285 TGTGCTAGCTTGAATTTGTTCCCGCGGATTTGAAGGGCTTGAAAGATAAAATTCCTTATT 1344  
Db 311 TGTGCTAGCTTGAATTTGTTCCCGCGGATTTGAAGGGCTTGAAAGATAAAATTCCTTATT 370  
Qy 1345 TTCAAGAGCTTGGTTGACTTATCTGACCTGATGCGCTGTTTAAATGCCCTCAAGGCA 1404  
Db 371 TTCAAGAGCTTGGTTGACTTATCTGACCTGATGCGCTGTTTAAATGCCCTCAAGGCA 430  
Qy 1405 AAAGCGAGCGCGGTATCGGTGACAGCTACCGGATGTCAATCGGCACTGGGCAAA 1464  
Db 431 AAAGCGAGCGCGGTATCGGTGACAGCTACCGGATGTCAATCGGCACTGGGCAAA 490  
Qy 1465 TAGCGACTTGGCGAAGTCAATTGCTGCTGCTGCAAGAGCGCGCATTTCCGCGTGTGTCG 1524  
Db 491 TAGCGACTTGGCGAAGTCAATTGCTGCTGCTGCAAGAGCGCGCATTTCCGCGTGTGTCG 550  
Qy 1525 ATTTTATCTTCAACACACCTCCACGACACGATGAGGCGCAAGCTGCGCGCGCGCG 1584  
Db 551 ATTTTATCTTCAACACACCTCCACGACACGATGAGGCGCAAGCTGCGCGCGCGCG 610  
Qy 1585 ACCCGCTTTTCGCAATTTCTATATATTTTCCCGAGCGCGGATGCCGACCAATACG 1644  
Db 611 ACCCGCTTTTCGCAATTTCTACTATATATTTTCCCGAGCGCGGATGCCGACCAATACG 670  
Qy 1645 ACCGACCGCTGGCGAATCTTCCCGACGACGACCGCGGCGCTTCTGCGCACTGGAAG 1704  
Db 671 ACCGACCGCTGGCGAATCTTCCCGACGACGACCGCGGCGGCTTCTGCGCACTGGAAG 730  
Qy 1705 ACGGACGCTGGGTGTGGAGACCTTCAATTTCTTCCCAATGGGACTTGAATACAGCAACC 1764  
Db 731 ACGGACGCTGGGTGTGGAGACCTTCAATTTCTTCCCAATGGGACTTGAATACAGCAACC 790  
Qy 1765 CGTGGGTATTTCGCGGATGGCGGGGAAATCTGTTCTGTCGAACTTGGCGTGTGACA 1824  
Db 791 CGTGGGTATTTCGCGGATGGCGGGGAAATCTGTTCTTGTGCCAACTTGGCGTGTGACA 850  
Qy 1825 TCCTCGGTATGATGCGGTGTGCTTTATTTTGAAAAAATAATGGGGACAACTGCGAAAAACC 1884

RESULT 7

851 TCCTGCTATGATGCGGTTCCTTTATTTGAAAAAATAATGGGGACAAAGCTCGAAAAACC 910  
1885 TGCCGCGAGCGCACGCCCTCATCGCGCGCTTCAATGCGGTATGCTATTATGCGCGCGCG 1944  
911 TGCCGCGAGCGCACGCCCTCATCGCGCGCTTCAATGCGGTATGCTATTATGCGCGCGCG 970  
1945 CCGTGTCTTCAAAATCCGAAGCCATCGTCCACCCCGACCAAGTGTCCCAATATATCGGGC 2004  
971 CCGTGTCTTCAAAATCCGAAGCCATCGTCCACCCCGACCAAGTGTCCCAATATATCGGGC 1030  
2005 AGAAGCAATGCCAAATCGGTTACAACCCCTGCAAAATGGCATTTGTTGTGAAACACCCCTTG 2064  
1031 AGAAGCAATGCCAAATCGGTTACAACCCCTGCAAAATGGCATTTGTTGTGAAACACCCCTTG 1090  
2065 CCACGCGCGAAGTCAACCTGCTCTCATCAGCGCTGACCTACCGCCCAACACCTGCCCGAGC 2124  
1091 CCACGCGCGAAGTCAACCTGCTCTCATCAGCGCTGACCTACCGCCCAACACCTGCCCGAGC 1150  
2125 ATACCGCTGGTCAACTAGTCCGAGCGCAACGACGACATCGGCTGACGCTTTGCCGATG 2184  
1151 ATACCGCTGGTCAACTAGTCCGAGCGCAACGACGACATCGGCTGACGCTTTGCCGATG 1210  
2185 AAGAAGCGGCATATCTGGGCAATAGCGGCTACGACCAACCGCCCAATTCCTCAACCGCTTCT 2244  
1211 AAGAAGCGGCATATCTGGGCAATAGCGGCTACGACCAACCGCCCAATTCCTCAACCGCTTCT 1270  
2245 TCGTCAACCGTTTCGACGCGGAGCTTGCCTGCTGGGTACCGTTTCCAAATCAACCCCAAGCA 2304  
1271 TCGTCAACCGTTTCGACGCGGAGCTTGCCTGCTGGGTACCGTTTCCAAATCAACCCCAAGCA 1330  
2305 CAGCGCACTCCCGTGTGCTGCTGAGTACAGCGCGGCTTGGTGGCTTGGCGCAAGACGATC 2364  
1331 CAGCGCACTCCCGTGTGCTGCTGAGTACAGCGCGGCTTGGTGGCTTGGCGCAAGACGATC 1390  
2365 CCACGCGGTGACCGGCATCAAACTCTGTGACAGCATTTGCTTTGAGTACCGCGGTCTGC 2424  
1391 CCACGCGGTGACCGGCATCAAACTCTGTGACAGCATTTGCTTTGAGTACCGCGGTCTGC 1450  
2425 CGCTGATTTACCTAGGCGAGCAAGTGGTACGCTCAATGACGACGACTGCTGCGCAAGCA 2484  
1451 CGCTGATTTACCTAGGCGAGCAAGTGGTACGCTCAATGACGACGACTGCTGCGCAAGCA 1510  
2485 GCAATAGACGACGACGCGGCTTGGCGCACCGTCCGCGCTACACGAAGCCCTGTACG 2544  
1511 GCAATAGACGACGACGCGGCTTGGCGCACCGTCCGCGCTACACGAAGCCCTGTACG 1570  
2545 GCAACGCAAGCATCCGTCGACGCGCGGCGCAATCTATCAGGCGCTTGGCGCATATGA 2604  
1571 GCAACGCAAGCATCCGTCGACGCGCGGCGCAATCTATCAGGCGCTTGGCGCATATGA 1630  
2605 TTGCGGTCCGCCAAAGCAATCCGCGCTTTCGACGCGCGGCGAGGCTGTTTACATTCAACACCA 2664  
1631 TTGCGGTCCGCCAAAGCAATCCGCGCTTTCGACGCGCGGCGAGGCTGTTTACATTCAACACCA 1690  
2665 ACAACAGCAATCATGCGGTACATCCGCAACAAATGCGCTTTTGGCATTCGTTAACTTCA 2724  
1691 ACAACAGCAATCATGCGGTACATCCGCAACAAATGCGCTTTTGGCATTCGTTAACTTCA 1750  
2725 GCGAATATCCGCAACCGTTTACCGCGCATACCTGCAAGCCATGCGCTTCAAGGCGCAAG 2784  
1751 GCGAATATCCGCAACCGTTTACCGCGCATACCTGCAAGCCATGCGCTTCAAGGCGCAAG 1810  
2785 ACCTCATCGGTGGCAAACTGTGACGCTGAATCAGGATTTGACGCTTTCAGGCGCTATCAGG 2844  
1811 ACCTCATCGGTGGCAAACTGTGACGCTGAATCAGGATTTGACGCTTTCAGGCGCTATCAGG 1870  
2845 TCATGCTGCTCGAATCGCTGACGCGCTTC 2877  
1871 TCATGCTGCTCGAATCGCTGACGCGCTTC 1903

AX107016	AX107016	6951 bp	DNA	linear	PAT 30-APR-2001
LOCUS	Sequence 7 from Patent WO0125449.				
DEFINITION	AX107016				
ACCESSION	AX107016				
VERSION	AX107016.1	GI:13922550			
KEYWORDS	synthetic construct				
SOURCE	synthetic construct				
ORGANISM	artificial sequences.				
REFERENCE	1				
AUTHORS	Bergs, H., Polakowski, T., Held, A. and Gallert, K.C.				
TITLE	Novel immobilizable amylosucrase, the use thereof and method for producing poly(1,4-alpha-glucan)				
JOURNAL	Patent: WO 0:25449-A 7 12-APR-2001;				
	Axiva GmbH (DE)				
FEATURES	Location/Qualifiers				
source	1..6951				
	/organism="synthetic construct"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:32630"				
	/notes="Beschreibung der kunstlichen				
	Sequenz: Expressionsvektor pGEX-4P-1-AmSu5"				
BASE COUNT	1658 a 1791 c 1759 g 1633 t				
ORIGIN					
	Query Match 64.9%; Score 1891.4; DB 6; Length 6951;				
	Best Local Similarity 99.9%; Pred. No. 0;				
	Matches 1892; Conservative 0; Mismatches 1; Indels 0; Gaps 0;				
QY	985 TGATTTTACAGTACCTCAAAACCGCATCTTGGACATCTACGCCCGACACGGCGCG 1044				
Db	946 TGATTTTACAGTACCTCAAAACCGCATCTTGGACATCTACGCCCGACACGGCGCG 1005				
QY	1045 GCATCGAAAAATCGGAAGACTGGCGCGCAGTTTTCGCGCGCGATGGATACGCATTTCGCCA 1104				
Db	1006 GCATCGAAAAATCGGAAGACTGGCGCGCAGTTTTCGCGCGCGATGGATACGCATTTCGCCA 1065				
QY	1105 AACTGATGAACGAACTCGACAGCGTGTTACGGCAACAGACGAGCCCTGCTGCCTATGCTGG 1164				
Db	1066 AACTGATGAACGAACTCGACAGCGTGTTACGGCAACAGACGAGCCCTGCTGCCTATGCTGG 1125				
QY	1165 AAATGCTGCTGGCGCAGGCATGGCAAAAGCTATTTCCTCAACGCAACTCATCTCTAAAAGATA 1224				
Db	1126 AAATGCTGCTGGCGCAGGCATGGCAAAAGCTATTTCCTCAACGCAACTCATCTCTAAAAGATA 1185				
QY	1225 TCGATATCGCGCGGGAANAACACCCCGATTGGATTTTGTCCACAAACAAGTCGGCGCGCG 1284				
Db	1186 TCGATATCGCGCGGGAANAACACCCCGATTGGATTTTGTCCACAAACAAGTCGGCGCGCG 1245				
QY	1285 TGTCTACGTTGATTGTTGTTGCGCGCGATTGAAGCGCTTGAAGAGATAAAATTCCTTATT 1344				
Db	1246 TGTCTACGTTGATTGTTGTTGCGCGCGATTGAAGCGCTTGAAGAGATAAAATTCCTTATT 1305				
QY	1345 TTCAAGAGCTTGTTTGAAGTATCTGCACCTGATGCGCGCTGTTTAAATGCGCTGAAGGCA 1404				
Db	1306 TTCAAGAGCTTGTTTGAAGTATCTGCACCTGATGCGCGCTGTTTAAATGCGCTGAAGGCA 1365				
QY	1405 AAACGACGCGCGCTATGCGGTGACGAGCTACCGCGATGTCAATTCGGGCACTGGGCAAA 1464				
Db	1366 AAACGACGCGCGCTATGCGGTGACGAGCTACCGCGATGTCAATTCGGGCACTGGGCAAA 1425				
QY	1465 TAGCGCACTTGCAGAGTCAATTGCTCGGTGTCACGAAGCGCGCATTTTCGCGCGTGTGTCG 1524				
Db	1426 TAGCGCACTTGCAGAGTCAATTGCTCGGTGTCACGAAGCGCGCATTTTCGCGCGTGTGTCG 1485				
QY	1525 ATTTTATCTTCAACACCACTCCAAACGAACAGAAATGCGGCGCAACGCTGGCGCGCGCGCG 1584				
Db	1486 ATTTTATCTTCAACACCACTCCAAACGAACAGAAATGCGGCGCAACGCTGGCGCGCGCGCG 1545				
QY	1585 ACCCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGCGAATGCCCGACCAATACG 1644				
Db	1546 ACCCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGCGAATGCCCGACCAATACG 1605				



```

Db      2686 GCGAATATCCGCAACCGTACCGGCGCATACCGCTCAAGCCATGCGCTTCAAGCGCGACG 2745
QY      2785 ACCTCATCGTGGCGCAAACTGTGAGCTGAACTGAGATTGACGCTTCAGCGCTATCAGG 2844
Db      2746 ACCTCATCGTGGCGCAAACTGTGAGCTGAACTGAGATTGAGCTTCAGCCCTATCAGG 2805
QY      2845 TCATGTGCTGGAATGCGCTGACGAGCTTTC 2877
Db      2806 TCATGTGCTGGAATGCGCTGACGAGCTTC 2838

RESULT 8
LOCUS   AY099334                1967 bp    DNA        linear    BCT 13-JAN-2003
DEFINITION   Neisseria meningitidis strain 93246 amylosucrase Ams (ams) gene,
              complete cds.
ACCESSION   AY099334
VERSION     AY099334.1   GI:27728141
KEYWORDS    .
SOURCE      Neisseria meningitidis
            Neisseria meningitidis
            Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
            Neisseriaceae; Neisseria.
REFERENCE   1 (bases 1 to 1967)
            Zhu, P., Teang, R.S.W. and Tsai, C.M.
            Nonencapsulated Neisseria meningitidis Strain Produces Amylopectin
            from Sucrose: Altering the Concept for Differentiation between N.
            meningitidis and N. polysacchara
            J. Clin. Microbiol. 41 (1), 273-278 (2003)
REFERENCE   2 (bases 1 to 1967)
            Zhu, P.
            Direct Submission
            Submitted (23-APR-2002) LBP/DBPAP, CBER, 8800 Rockville Pike,
            Bethesda, MD 20892, USA
            Location/Qualifiers
              1..1967
               /organism="Neisseria meningitidis"
               /mol_type="genomic DNA"
               /strain="93246"
               /db_xref="taxon:487"
              21..1931
               /gene="ams"
              21..1931
               /gene="ams"
               /codon_start=1
               /transl_table=11
               /product="amylosucrase Ams"
               /protein_id="AA051152.1"
               /db_xref="GI:27728142"
               /translation="MLTPTQVGLILQVLTIRLIDIVTPEORAGIEKSEDWROPSRRM
               DTHPKLNEIDSVYGNNEALLPMLLELAQWOSYORSASLKNIDIERENYPDWIL
               SNKQVGVYVDLFDGLKGLKDIHYEQLGLTYLHLMPLPKPEKSDGAYVSSY
               RDVNPALGTIGDLREIVIALREAGISAVVDIEFNHTSNEHWAQRCAAGDPLFNPYY
               IFPDRMPQVDRTLEIFPDHGGGSLSDGRWWTTFNSFDNLNYPWFRRM
               AGEMLPLAGVDIRMDVAIFIKWQMTSCENLPQAHALIRAFNAVIRIAAPVFFK
               SEAIIVHPQVQVYIGQDEQCIQNPQLQWALLNTLATREVNLLHQALTYRNLPEHTA
               WUNYVRSHDDIGTWPADDAAYLGIISYDHRQFLNRPVNRPDGSEFARGVPFOYNPT
               GDCRVSGTSAALAQNDPAVSRILKLYSIALSTGLPLIYLGDVETPDNDGWQ
               DSNKDDSRWAHRPYNALYAQRNDPSTAAGIQGLRMIAYRQSNRPFDGGRLYT
               FNNKHLIGTIRNALLAFNFSYPTVHTILQAFKHLHIGKIVSLNQDLIT
               LQPYQWMLLEIA"
BASE COUNT      463 a      526 c      490 g      388 t
ORIGIN
Query Match      63.6%; Score 1853.4; DB 1; Length 1967;
Best Local Similarity 96.4%; Pred. No. 0;
Matches 1896; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
QY      937 TCAGAATCGGAGCGGCGATCATGTTGACCCCGCCACGACCAAGTCGGTTGATTTTACAGT 996
Db      1 TCAGAATCGGAGCGGCGATCATGTTGACCCCGCCACGACCAAGTCGGTTGATTTTACAGT 60

```

```

QY      997 ACCTCAAAACACGCAATCTTGACATCTACAGCCCGCAACAGCGCGCGGCATCGAANAAT 1056
Db      61 ACCTCAAAACACGCAATCTTGACATCTACAGCCCGCAACAGCGCGCGGCATCGAANAAT 120
QY      1057 CCGAAGACTGCGGCGAGTTTTCGCGCGCATGATGATGCAATTTCCCAAACTGATGACG 1116
Db      121 CCGAAGACTGCGGCGAGTTTTCGCGCGCATGATGATGCAATTTCCCAAACTGATGACG 180
QY      1117 AACTCGACAGCGTGTACGGCAACAAACGAAAGCCCTGCTGCTATCTGTAATGCTGCTGG 1176
Db      181 AACTCGACAGCGTGTACGGCAACAAACGAAAGCCCTGCTGCTATCTGTAATGCTGCTGG 240
QY      1177 CCGAAGACTGCGGCGAGTTTTCGCGCGCATGATGATGCAATTTCCCAAACTGATGACG 1236
Db      241 CCGAAGACTGCGGCGAGTTTTCGCGCGCATGATGATGCAATTTCCCAAACTGATGACG 300
QY      1237 CCGAAGACTGCGGCGAGTTTTCGCGCGCATGATGATGCAATTTCCCAAACTGATGACG 1296
Db      301 CCGAAGACTGCGGCGAGTTTTCGCGCGCATGATGATGCAATTTCCCAAACTGATGACG 360
QY      1297 ATTTGTTTGGCGGCGATTTGAAAGGCTGTGAAAGATGAAATTCCTTATTTTCAAGAGCTTG 1356
Db      361 ATTTGTTTGGCGGCGATTTGAAAGGCTGTGAAAGATGAAATTCCTTATTTTCAAGAGCTTG 420
QY      1357 GTTTGACTTATCTGACGCTGATGCGCTGTTTAAATGCCCTGAAAGCAAAATTCATTTTCAAGAGCTTG 1416
Db      421 GTTTGACTTATCTGACGCTGATGCGCTGTTTAAATGCCCTGAAAGCAAAATTCATTTTCAAGAGCTTG 480
QY      1417 GCTATGCGGTGACGAGCTACCGGATGTCAATCGCGCACTGGGCAATAGGCGAGCTTGC 1476
Db      481 GCTATGCGGTGACGAGCTACCGGATGTCAATCGCGCACTGGGCAATAGGCGAGCTTGC 540
QY      1477 GCGAAGTCAATGCTGCGCTGACGAGCGGCAATTCGCGCTGCTGATTTTATCTTCA 1536
Db      541 GCGAAGTCAATGCTGCGCTGACGAGCGGCAATTCGCGCTGCTGATTTTATCTTCA 600
QY      1537 ACCACACTCTCCACGAAACAGCAATGGGCGCAACGCTGCGCGCGCGCGGCAACCGCTTTTCG 1596
Db      601 ACCACACTCTCCACGAAACAGCAATGGGCGCAACGCTGCGCGCGCGCGGCAACCGCTTTTCG 660
QY      1597 ACAATTTCTACTATATTTTCCCGACCGCGGATGCGCGGCAATAGCGACCGGCGGCGGCGG 1656
Db      661 ACAATTTCTACTATATTTTCCCGACCGCGGATGCGCGGCAATAGCGACCGGCGGCGGCGG 720
QY      1657 GCGAATCTTCCCGGACGAGCACCGCGGCGGCTTCTCGCAACTGGAAGACGAGCGCTGGG 1716
Db      721 GCGAATCTTCCCGGACGAGCACCGCGGCGGCTTCTCGCAACTGGAAGACGAGCGCTGGG 780
QY      1717 TGTGAGAGACTTCAATTTCTTCCATGGAATTTGAAATACAGCAACCGCTGGGTATTC 1776
Db      781 TGTGAGAGAGCTTCAATTTCTTCCATGGAATTTGAAATTTACAGCAACCGCTGGGTATTC 840
QY      1777 GCGCAATGCGGCGGGAATGCTGTTCTTCCCACTTGGGCGTTGACATCTCTCGGTATGG 1836
Db      841 GCGCAATGCGGCGGGAATGATGTTCTTCCCACTTGGGCGTTGACATCTCTCGGTATGG 900
QY      1837 ATGCGGTGCGCTTTTATTTGGAACAAATGGGCAACGCTGCGAAGAACTGCGCGGCGGCG 1896
Db      901 ATGCGGTGCGCTTTTATTTGGAACAAATGGGCAACGCTGCGAAGAACTGCGCGGCGGCG 960
QY      1897 ACSCCTCATCTCCGCGCTTCAATGCGGTTATGGTATTTGGGCGGCGCGCGGCTGTTCTTCA 1956
Db      961 ACSCCTCATCTCCGCGCTTCAATGCGGTTATGGTATTTGGGCGGCGCGCGGCTGTTCTTCA 1020
QY      1957 AATTCGAAGCGCATCGTCCACCGCGGCAAGTCTGTCCTCAATATCATCGGCGGAGCAAGTCC 2016
Db      1021 AATTCGAAGCGCATCGTCCACCGCGGCAAGTCTGTCCTCAATATCATCGGCGGAGCAAGTCC 1080
QY      2017 AATTCGGTTACACCCCGCTGAAATGCGATGTTGGAACACCCCTGCGGCGGCGGAG 2076
Db      1081 AATTCGGTTACACCCCGCTGAAATGCGATGTTGGAACACCCCTGCGGCGGCGGAG 1140

```



Qy	2077	TCAACTGCTCCATTCAGGCGGTGACTACCGCCACAACCTTGCCTCCGAGCATACCGCTGGG	21136
Db	1141		
		TCAACTGCTCCATTCAGGCGGTGACTACCGCCACAACCTTGCCTCCGAGCATACCGCTGGG	1200
Qy	2137	TCAACTACGTCGCGCAGCCACGACGACATCGCTGAGCGTTTGGCGATGAAGACGCGGCAT	2196
Db	1201		
		TCAACTACGTCGCGCAGCCACGACGACATCGCTGAGCGTTTGGCGATGAAGACGCGGCAT	1260
Qy	2197	ATCTGGGCATAAAGCGGCTACGACCAACCGCAATTCCTCAACCGGTTCTTCGTCAACCGTTT	2256
Db	1261		
		ATCTGGGCATAAAGCGGCTACGACCAACCGGCAATTCCTCAACCGGTTCTTCGTCAACCGTTT	1320
Qy	2257	TGCACGCGAGCTTCGCTCTGGGGGTACCGTTCCAAATACACCCCAAGCACACGAGCGACTGCC	2316
Db	1321		
		TGCACGCGAGCTTCGCTCTGGGGGTACCGTTCCAAATACACCCCAAGCACACGAGCGACTGCC	1380
Qy	2317	GTGTCAGTGTTACAGCCGCGGCANTTGTTCGCTTGGCGCAAGCAGATCCCCACGCGCGTTG	2376
Db	1381		
		GGTTCAGCGGTACGTCGCGCGCATTTGCGAGGCTTGGTCAAAATGACCCCCACGCGGTCA	1440
Qy	2377	ACCGCATCAAACTCTTGTAACGACNTGCTTTTGAGTACCGGGGTCTCTCGCTCATTTACC	2436
Db	1441		
		CGCGCATCAAACTTTGTACAGCATCGATTTGAGTACCGCGGGCTCTCCCTCTCATTTACC	1500
Qy	2437	TAGCGCAGAGTGGGTACGCTCAATCAGACGACGACTGCTCGCAAGCAGCAATAAGAGCG	2496
Db	1501		
		TGCGCGCAGAGTTCGCGACGCCCAACGACGCGCTGGTGTGAGACGAGCAACAAAGCG	1560
Qy	2497	ACGACAGCGTGTGGGCGCACGCTCCGGCGCTACAAAGAGCCCTGTACGCGCAACGCAAGC	2556
Db	1561		
		ACGACAGCGTGTGGGCGCACGCTCCGGCGCTATTAACGAAGCCCTGTACGCGCAACGCAAGC	1620
Qy	2557	ATCGTGCACGCGACCGGGCAATCTATCAGGGCTTGGCGCATATGATTTGCGTCTCGCC	2616
Db	1621		
		ATCGTGCACGCGACCGGGCAATCTATCAGGGCTTGGCGCATATGATTTGCGTCTCGCC	1680
Qy	2617	AAAGCAATTCGCGCTTTCGACGGCGGCGAGGTGGTTACATTCACACCAACAAACAAGCACA	2676
Db	1681		
		AAAGCAATTCGCGCTTTCGACGGCGGCGAGGTGGTTACATTCACACCAACAAACAAGCACA	1740
Qy	2677	TCATCGGCTACATCCGCAACAATCGCTTTTGGCAATTCGGTAACTTCAGCGAATATCCGC	2736
Db	1741		
		TCATCGGCTACATCCGCAACAACCGGCTTTTGGCAATTCGGCACTTCAGCGAATATCCGC	1800
Qy	2737	AAAACGGTTACCGGCATACCTGCAAGCCATGCCCTTCAGGGCGCAGCTCATCGGTG	2796
Db	1801		
		AAAACGGTTACCGGCATACCTGCAAGCCATGCCCTTCAGGGCGCAGCTCATCGGTG	1860
Qy	2797	GCAAAACTGTCAAGCTCAATCAGGATTTGACGCTTTCAGCCCTATCAGGTCATGTGGCTCG	2856
Db	1861		
		GTAANACTGTCAAGCTCAATCAGGATTTGACGCTTTCAGCCCTATCAGGTCATGTGGCTCG	1920
Qy	2857	AAATCGCTGACGACGCTTCCCAAAATGCGCTCTGAACCGTTTCAGA	2903
Db	1921	AAAACGGCTGACGACGCTTCCGGAATGCGCTCTGAACCGTTTCAGA	1967

RESULT 9	
AY099335	AY099335
LOCUS	Neisseria polysaccharea strain 85322 amyloucrase Ams (ams) gene, complete cds.
DEFINITION	Neisseria polysaccharea strain 85322 amyloucrase Ams (ams) gene, complete cds.
ACCESSION	AY099335
VERSION	AY099335.1 GI:27728139
KEYWORDS	
SOURCE	Neisseria polysaccharea
ORGANISM	Neisseria polysaccharea
	Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria.
REFERENCE	1 (bases 1 to 1967)
AUTHORS	Zhu, P., Tsang, R.S.W. and Tsai, C.M.
TITLE	Nonencapsulated Neisseria meningitidis Strain Produces Amylopectin from Sucrose: Altering the Concept for Differentiation between N.

JOURNAL	PUBLISHED	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES	source
meningitidis and N. polysaccharea	J. Clin. Microbiol. 41 (1), 273-278 (2003)	12517860	Zhu, P.	Direct Submission	Submitted (23-APR-2002) LBP/DBAP, CBER, 8800 Rockville Pike, Bethesda, MD 20892, USA	Location/Qualifiers	1..1967
							/organism="Neisseria polysaccharea"
							/mol_type="genomic DNA"
							/strain="85322"
							/db_xref="taxon:489"
							21..1931
							/gene="ams"
							21..1931
							/gene="ams"
							/codon_start=1
							/transl_table=11
							/product="amylosucrase Ams"
							/protein_id="AAM51153.1"
							/db_xref="GI:2728140"
							/translation="MULTPQQVGLILQHLKTRILDIYTPQRAGIEKSHDWQFSRRM
							DNFKPKMELVDSVYGNNAALFLPMLLELLAQWQSORSASLKNIDIRENNPDMIL
							SKHQVGVGGLVDFADGLKGLDKIPYQELAGLTYLHLMLPKCPKGGDGGYAVSSY
							RVNPLGTTIDDLREYIAALHBAIGASVAVDFINHTSNEHEWAQRCAAGDPDGNFY
							IPEDRPQYQRTLEIPDPDHGGFSOLEDRWVWTFNSFQMDLNTSNWVVFRAK
							AGEMFLANGLVDILNDVAFAIWKMGTCENLQAHALIAEFNAVMIAAPAVPEFK
							SEAIHPDQVQVIGDQCOIYNPLQMLNNTLATREVNILLHQALTVRHNLPEHTS
							WNYVRSHDIDGWTFADEDAAYIGISYDHRQFLNRPFFVNRFDGSGFARGVPQYINPT
							GDRVSTSAALAGLQNDPHAVSRILKJLSYALSTGGPLIYLGDEVGTDPNDGWAQ
							DGNSDSSRWAAHPRVNEALYQRNDPTAAGQIYQLRHMTAVRQSNPRFPDGRGLV
							FNTNNKHIIGYIRNNALLAFGNFSEYPQTVTAHTLQAMPFKAHDLGGKTVSLNQDLT
							LQPYQVMLEIA"
BASE COUNT	459 a	631 c	489 g	388 t			
ORIGIN							
Query Match	63.4%;	Score 1847;	DB 1;	Length 1967;			
Best Local Similarity	96.2%;	Pred. No. 0;					
Matches 1892;	Conservative	0;	Mismatches	75;	Indels	0;	Gaps 0;
Qy	937	TCAGATCGGAGCAGGCATCATGTTGACCCCAACGACGCAAGTCGGTTTGATTTTACAGT	996				
Db	1	TCAGATCGGAGCAGGCATTTATGTTGACCCCAACGACGCAAGTCGGTTTGATTTTACAGC	60				
Qy	997	ACCTCAAAACACGCATCTTGGACATCTACACGCCCAACAGCGCGCGGCATCGAAAAAT	1056				
Db	61	ACCTCAAAACGCGCATCTTGGACATCTACACGCCCAACAGCGCGCGGCATCGAAAAAT	120				
Qy	1057	CCGAAGACTCGCGCGAGTTTTTCGGCGGCAATGATACGATTCGCCCAACTCATGACG	1116				
Db	121	CCGAAGACTCGCGCGCATTTTTCGGCGGATGATACGATTCGCCCAATTCATGACG	180				
Qy	1117	AACTCGACACGGTGTACGGCAACAAAGAACCCCTGCTGCTATGCTGGAAATGCTGCTGG	1176				
Db	181	AACTCGACACGGTGTACGGCAACAAAGAACCCCTGCTGCTATGCTGGAAATGCTGCTGG	240				
Qy	1177	CGCAGGCATCGCAAGCATTTCCCAACGCACTCATCTCTTTAAAGATATCGATATCGCGC	1236				
Db	241	CGCAGGCATCGCAGAGCTATTTCCTCAACGCGCGCTCCCTCGAAAAACATTTGACATCGAAC	300				
Qy	1237	CGCAACACACCCGATTTGATTTGTTGCCAACAAACAAAGTCGCGCGCTGCTGCTGCTTG	1296				
Db	301	CGCAAAACACCCCGATTGATTTTGTCCAAACAAACAAAGTCGCGCGCTGCTGCTGCTTG	360				
Qy	1297	ATTTGTTGTCGGCGGATTTGAAGCGCTTGAAAGATATAAAATTCCTTATTTTCAAGAGCTTG	1356				
Db	361	ATTTGTTGTCGGCGGATTTGAAGCGCTTGAAAGATATAAAATTCCTTATTTTCAAGAGCTTG	420				
Qy	1357	GTTTGACTTATCTGCACCTGATCGCGCTGTTTAAATGCTCCTGTAAGCAAGCGCGCGG	1416				
Db	421	GTTTGACTTATCTGCACCTGATCGCGCTGTTTAAATGCTCCTGTAAGCAAGCGCGCGG	480				

Qy	1417	GCTATGGGTT	CAGCAGCT	TACCGCAT	GTCAAT	TCGGCA	CTGGGC	ACAATAGG	CGACTTGC	1476
Db	481	GCTATGCGGT	CAGCAGCT	TACCGCAT	GTCAAT	TCGGCA	CTGGGC	ACAATAGG	CGACTTGC	540
Qy	1477	GCGAAGTCA	TTGCGCTG	CAAGAG	CGCGCAT	TTTCCG	CGCTCGT	CGATTTAT	CTTCTCA	1536
Db	541	GCGAAGTCA	TTGCGCGT	GCACAG	AGCGGT	ATTTCC	CGCGTGT	CGATTTAT	CTTCTCA	600
Qy	1537	ACCACACT	CCAAAGAA	CACGAAT	TGGGCG	CAAA	CGCTGCG	CGCGCG	CGACCGCTTTTCG	1596
Db	601	ACCACACT	CCAAAGAA	CACGAAT	TGGGCG	CAAG	CTGCGCG	CGCGCG	CGACCGCTTTTCG	660
Qy	1597	ACAAATTT	CTACTAT	TATTTT	CCCGAC	CGCGCG	ATGCTCC	CGACCA	ATACGACCGACCTTGC	1656
Db	661	ACAAATTT	CTACTAT	TATTTT	CCCGAC	CGCGCG	ATGCTCC	CGACCA	ATACGACCGACCGCTTGC	720
Qy	1657	GCGAAATC	TTCTCCG	CACGAC	CGCGCG	CGCTTCT	CGCAACT	TGGAAG	ACGACGCTTGGG	1716
Db	721	GCGAAATC	TTCTCCG	CACGAC	CGCGCG	CGCTTCT	CGCAACT	TGGAAG	ACGACGCTTGGG	780
Qy	1717	TGTGGA	CGACTT	CAATTC	TTTCA	ATGGGA	CTTGAAT	TACAGAA	CCCGTGGGTATTC	1776
Db	781	TGTGGA	CGACGT	TTCAAT	TTCTT	CAATGG	CACTTGAAT	TACAGAA	CCCGTGGGTATTC	840
Qy	1777	GCGCAAT	TGGGCG	GAATCT	GTCTTC	TCCCACT	TGCGCA	ACTTGGCG	TGACATCCTCGGTATGG	1836
Db	841	GCGCAAT	TGGGCG	GAATCT	GTCTTC	TCCCACT	TGCGCA	ACTTGGCG	TGATCTTGGGTATGG	900
Qy	1837	ATGCGG	TTCGCT	TTATTT	TGAAAA	CAAAAT	TGGGGA	CAAGCT	TCGAAAAA	1896
Db	901	ATGCGG	TTCGCT	TTATTT	TGAAAA	CAAAAT	TGGGGA	CAAGCT	TCGAAAAA	960
Qy	1897	ACGCCCT	CAATCC	CGCG	CGTTCA	ATCCG	GTATG	GTATG	TCGGCG	1956
Db	961	ACGCCCT	CAATCC	CGCG	CGTTCA	ATATG	CGCTT	ATG	GTATG	1020
Qy	1957	AATCCGA	AGCAT	CGTCT	CACCC	CAACAA	GTGTC	CAATAT	CATCGG	2016
Db	1021	AATCCGA	AGCAT	CGTCT	CACCC	CAACAA	GTGTC	CAATAT	CATCGG	1080
Qy	2017	AAATCG	GTATCA	ACCCCT	CTGCA	AAATG	GCATTT	GTGTG	GAACAC	2076
Db	1081	AAATCG	GTATCA	ACCCCT	CTGCA	AAATG	GCATTT	GTGTG	GAACAC	1140
Qy	2077	TCACCT	GTCTC	ATC	AGCG	CTGAC	CTAC	CGCC	CAACCT	2136
Db	1141	TCACCT	GTCTC	ATC	AGCG	CTGAC	CTAC	CGCC	CAACCT	1200
Qy	2137	TCAACT	AGTCT	CGCAG	CCACA	CGACAT	CGCTG	GCAG	CTTTG	2196
Db	1201	TCAACT	AGTCT	CGCAG	CCACA	CGACAT	CGCTG	GCAG	CTTTG	1260
Qy	2197	ATCTGG	GCATAG	AGCG	GGCTAC	GCAC	CGCC	CAATTC	CTCAAC	2256
Db	1261	ATCTGG	GCATAG	AGCG	GGCTAC	GCAT	CACCG	CAATTC	CTCAAC	1320
Qy	2257	TCGAC	GCAGCT	CTCG	CTG	TGGGTAC	CGTTT	CCAATAC	AAACCA	2316
Db	1321	TCGAC	GCAGCT	CTCG	CTG	TGGGTAC	CGTTT	CCAATAC	AAACCA	1380
Qy	2317	GTGTC	AGTGF	PAC	AGCG	CGGCA	ATTG	TGGCG	CAAG	2376
Db	1381	GCGT	CAGCG	GTAC	GTCCG	CGCG	CACTT	GGCTT	GGCTT	1440
Qy	2377	ACCG	CAATCA	ACTCT	GTCT	PAC	AGCA	TTGCTT	GAGT	2436
Db	1441	GCGGC	ATCA	AACTTT	GTGT	ACH	AGCAT	CGCAT	TTGAGT	1500
Qy	2437	TAGG	CGAC	GAAG	TGG	GTAC	GTCT	CAAT	GAC	2496
Db	1501	TCGG	CGAC	GAAG	TCGG	CA	CGCC	CAAC	CA	1560

Qy	2497	ACGACAGCGCTTGGGGCGCAACGGTCGGCGGTACAACAGAGCCCTGTGTACGGCGCAACGCAACG	2555
Db	1561	ACGACAGCGCTTGGGGCGCACCGTCGGCGGTACAACAGAGCCCTGTGTACGGCGCAACGCAACG	1620
Qy	2557	ATCCGTGCAGCGCAGCGCGGCGAAATCTATCAGGGCTTTGGGCCATATGATTGCCGTCCGCC	2816
Db	1621	ATCCGTGCAGCGCAGCGGCGAATCTATCAGGGCTTTGGGCCATATGATTGCCGTCCGCC	1680
Qy	2617	AAGCAATCCGCGCTTGCGAGCGCGCGAGGCTGGTTACTTAACACCAACAAACGACACA	2676
Db	1681	AAGCAATCCGCGCTTGCGAGCGCGCGAGGCTGGTTACTTTCAACACCAACAAACGACACA	1740
Qy	2677	TCAATCGGCTACATCCGCGCAACGAATGCGCTTTGGCATTCGGTAACCTCAGCGATATCCGC	2736
Db	1741	TCAATCGGCTACATCCGCGCAACGCGCTTTGGCATTCGCAACTTCAGCGATATCCGC	1800
Qy	2737	AAACCGTTAACCGGCGCATACCCTGCAGACCATGCCCTTCAAGGCGCAGCACTCATCGGTG	2796
Db	1801	AAACCGTTAACCGGCGCATACCCTGCAGACCATGCCCTTCAAGGCGCAGCACTCATCGGCG	1860
Qy	2797	GCAAACTGTACGCTGAATCAGAGTTAGCGCTTCAGGCCCTATCAGGTCATGTGGCTCG	2856
Db	1861	GCAAACTGTACGCTGAATCAGAGTTAGCGCTTCAGGCCCTATCAGGTCATGTGGCTCG	1920
Qy	2857	AAATCGGCTGACGCAACGCTTCCCATAATGCCGTCTGAACCGTTTCAGA	2903
Db	1921	AAATCGGCTGACGCAACGCTTCCCATAATGCCGTCTGAACCGTTTCAGA	1967

RESULT 10  
AX300255/c

LOCUS AX300255 Sequence 81 from Patent WO01.85772.  
DEFINITION AX300255  
ACCESSION AX300255  
VERSION AX300255.1 GI:17381646  
KEYWORDS  
SOURCE Neisseria meningitidis  
ORGANISM Neisseria meningitidis  
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
Neisseriaceae; Neisseria.

REFERENCE  
1 Tang,C.U.  
AUTHORS Virulence genes, proteins, and their use  
TITLE Patent: WO 0185772-A 81 15-NOV-2001;  
JOURNAL Microscience Limited (GB)  
FEATURES Location/Qualifiers  
source 1..795  
    /organism="Neisseria meningitidis"  
    /mol\_type="genomic DNA"  
    /db\_xref="taxon:487"

CDS  
1..>795  
    /notes="unnamed protein product"  
    /codon\_start=1  
    /transl\_table=11  
    /protein\_id="CAD12960.1"  
    /db\_xref="GI:17381647"  
    /translation="MAAFVSVCACMFCSLVVNTVIGSGRTKIAVLPLVARDAVLVA  
VLQIKNLPFDIVFRADELCAGSTGEVLRHTQTVDALPDKPPLLFTFRREGEGGSF  
PCSDDYVELLDALIISRLPDIIDIELPSGETAVRCANAOKNGIALICNHEPHRT  
PPQREIVCKLKWEDCGADICKIAMVQSAEDVLTLSATLKAELAAKPIVTMSMGQ  
TGAVSLACGVFGSSITFGSGTQNSAPQIGYSALRATLDDLENGAD"

BASE COUNT 166 a 238 c 223 g 168 t

ORIGIN

Query Match 22.5%; Score 655; DB 6; Length 795;  
Best Local Similarity 96.4%; Pred. No. 3.8e-127;  
Matches 670; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Qy	1	GAGTTTGGCTTCCGAAACGACGATGCTGTGACCGCAACACCTGTCCGCGACGGCGG	60
Db	728	GAGTTTGGCTTCCGAAACGACGATGCTGTGACCGCAACACCTGTCCGCGACGGCGG	669
Qy	61	CTGACGCCCGCTTTTGCCCCATCGACATCGTAACAAATCGGTTTTGTGGCAAGCTTTTC	120



```

/codon_start=1
/transl_table=11
/product="hypothetical protein NMA1443"
/protein_id="CAB84679.1"
/db_xref="GI:7380096"
/db_xref="SPTREMBL:OSJUR9"
/translation="MIYKMCNPKKCAENKMKRMVFIENGCSATVHRTGIWKLIH
IDIIIPFIA"
gene complement (4509. .6960)
/gene="NMA1444"
CDS complement (4509. .6960)
/gene="NMA1444"
/note="NMA1444, probable P-type cation-transporting
ATPase, len: 823 aa; similar to many e.g. SH:COXA_ENTHR
(EMBL:L13292), CopA, Enterococcus hirae
copper/potassium-transporting ATPase A (EC 3.6.1.36) (727
aa), fasta scores; E(): 0, 34.0% identity in 744 aa
overlap. Similar to NMA1539, fasta scores; E(): 0, 34.4%
identity in 735 aa overlap. Contains hydrophobic, probable
membrane-spanning regions. Contains two pfam matches to
entry PF00122 E1-E2 ATPase, E1-E2 ATPases and PF00154
E1-E2 ATPases phosphorylation site"
/codon_start=1
/transl_table=11
/product="putative P-type cation-transporting ATPase"
/protein_id="CAB84680.1"
/db_xref="GI:7380097"

Query Match      22.5%; Score 555; DB 1; Length 329861;
Best Local Similarity 96.4%; Pred. No. 4.1e-127;
Matches 670; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

1  GAGTTTTCGTTCCCGAACCGAAGCTGATGTTGAGCGCGAACACTGTCCGCGCAAGCGCG 60
243562 GAGTTTTCGTTCCCGAACCGAAGCTGATGTTGAGCGCGAACACTGTCCGCGCAAGCGCG 243562

61  CTGACCGCGCCCTTTTGCCCCCATCGACATCGTAACAATPCGGTTTGGTGGCAAGCTCTTTC 120
243502 CTGACCGCGCCCGCTGTGCCCATCGACATCGTAACAATPCGGTTTGGTGGCAAGCTCTTTC 243444

121  GCTTTGACGCTGGCAGAAAGCAAAAGTCAGACAGCTCTCCGCGCTTTCGGGCATCACCGCA 180
243442 GCTTTGACGCTGGCAGAAAGCAAAAGTCAGACAGCTCTCCGCGCTTTCGGGCATCACCGCA 243388

181  ATTTTGCAGATGTCGCGCGCGCAGTCTCTCCATCTGTTTTCAGACGGCATACGATTTCTCT 240
243382 ATTTTGCAGATGTCGCGCGCGCAGTCTCTCCATCTGTTTTCAGACGGCATACGATTTCTCT 243322

241  TCGCGCGCTGCGGTGAACTCAATGATTCAGACGAGCGCGCGCGATGCGCGTTTGTGTA 300
243322 TCGCGCGCTGCGGTGAACTCAATGATTCAGACGAGCGCGCGCGATGCGCGTTTGTGTA 243266

301  GCATTCGCGCAGCGCGCGCGAGCGCGGTTTTCGCCGAAATAAGCTCGATATCGATAATG 360
243262 GCATTCGCGCAGCGCGCGAGCGCGGTTTTCGCCGAAATAAGCTCGATATCGATAATG 243200

361  TCGCGCAGCGCGCTTTCATTCAGCGAGTCGACAGTTCAAAAATAATATCGTCGAAACAC 420
243202 TCGCGCAGCGCGCTTTCATTCAGCGAGTCGACAGTTCAAAAATAATATCGTCGAAACAC 243144

421  GGGAAACAGCGCGCTTCGCCATGCCCTCTGAAAGTAAACAGCGAGCTGTCTGGCGGAGC 480
243142 GGGAAACAGCGCGCTTTCGCCATGCCCTCTGAAAGTAAACAGCGAGCTGTCTGGCGGAGC 243088

481  GCGTTCGCGGAGCGGTCTGCGGTGTGGCGCAATACATTTTCGCCGATGTCGCCCGCGCAT 540
243082 GCGTTCGCGGAGCGGTCTGCGGTGTGGCGCAATACATTTTCGCCGATGTCGCCCGCGCAT 243026

541  AAATTCGCGCGGAACTCGACGATATCGAGCGCGAGTTTGTGATTTGGTCAAGTACGGCG 600
243022 AAATTCGCGCGGAACTCGACGATATCGAGCGCGAGTTTGTGATTTGGTCAAGTACGGCG 242966

601  GAAAGTACGGCGCATTCGCGCGGCAAGCGCGACGCGGATTTTGGTTCGCTTCGCTTCGCG 660

```

```

Db 242962 GAAAGTAAGCGGCGATCGCGGCGCAAGCGCAAGCGGATTTTGGTGGTCCGCTTCGG 242903
QY 661 ATAAAGCGTCTTTTACGCGTCAAGCGTGGTGTGCAT 695
Db 242902 ATAAAGCGTCTTTTACGCGTCAAGCGGCGGCAACAT 242868

RESULT 12
AX565901/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. .765
/organism="Neisseria meningitidis serogroup B"
/mol_type="genomic DNA"
/db_xref="taxon:491"
1..765
/notes="unnamed protein product; 3-dehydroquinate
dehydratase"
/codon_start=1
/transl_table=11
/protein_id="CAD58481.1"
/db_xref="GI:26001181"
/translation="MCSLVKNTVIGSGRTKIAVLVDAEALSAVLEQLKNMPEP
IAEFRAFDPECAGSIEGLIHHQTVTDALDPKLFYFRHGGSGSPDCSDYVRELL
DALIERLDPIDIEFSGTAVRCVANAQNGNIAALCNHFRHTPPQESIVRELK
QXEDGADCKIAVPOSAEDVLITSLAKELAAKPIVTNMGQTGAVSRLAGQV
FGSSITFGSGTQNSAPGQIGVSAIRATDCLENGAD"
BASE COUNT 167 a 230 c 216 g 152 t
ORIGIN

Query Match 21.9%; Score 639; DB 6; Length 765;
Best Local Similarity 95.0%; Pred. No. 8.9e-124;
Matches 660; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 1 GAGTTTGGCTTCCGAAACGAAACGATGATGCTTGGAGCGCAACACCTGTCCGCAAGCGG 60
Db 595 GAGTTTGGCTTCCGAAACGAAACGATGATGCTTGGAGCGCAACCTGTCCGCAAGCGG 636
QY 61 CTGACCGCCCTTTTCCGCAATCGACATCGTAAACATCGTTGGTGGCAAGCTCTTTC 120
Db 635 CTGACCGCCCTTTTCCGCAATCGACATCGTAAACATCGTTGGTGGCAAGCTCTTTC 576
QY 121 GCTTTGAGCGTGGCAGAAACGAAAGTACAGACGCTTCTCCGCGCTTTGGCGCATCACCGCA 180
Db 575 GCTTTGAGCGTGGCAGAAACGAAAGTACAGACATCTTCCGCGCTTTGGCGCATCACCGCA 516
QY 181 ATTTGAGAGTCCGCGCGGCGATCTCTCCATCTGTTTCAGCGGCATAGATTTCTTCT 240
Db 515 ATTTGAGAGTCCGCGCGGCGGATCTCTCCATCTGTTTCAGCGGCATAGATTTCTTCT 456
QY 241 TGGCGCGGCTGGGTGAACATCATGATTCAGAGCGGCGGCGATGCGGTTTTTTTGA 300
Db 455 TGGCGCGGCTGGGTGAACATCATGATTCAGAGCGGCGGCGGATGCGGTTTTTTTGA 396
QY 301 GCATGCGCCACGCGCGCGCGGAGCGGCTTTCGCGGAAAGGCTCGATATCATGATAATG 360
Db 395 GCATTTGCCACGCGCGCGCGGAGCGGCTTTCGCGGAAAGGCTCGATATCATGATAATG 336

```

```

QY 361 TCGGCGAGGCGGCTTTCAATCAGCGAGTTCGAGCAGTTTCAAAATAAATAATCTCCGACAC 420
Db 335 TCGGCGAGGCGGCTTTCAATCAGCGAGTTCGAGCAGTTTCAAAATAAATAATCTCCGACAC 276
QY 421 GGGAAACGAGCGGCTTCCCATGCGGCTCTGACGTAAACAGCAGCGGCTTTCGCGGAGC 480
Db 275 GGGAAACGAGCGGCTTCCCATGCGGCTCTGACGTAAACAGCAGCGGCTTTCGCGGAGC 216
QY 481 CGCTGCGCGAGCGTCTCGGTGCGGCAATACTTTCGCGGATGCTGCGCGCGCATTTCCAAA 540
Db 215 CGCTGCGCGAGCGTCTCGGTGCGGCAATACTTTCGCGGATGCTGCGCGCGCATTTCCAAA 156
QY 541 AAATCGGCGGGAACCTCGACATATCGAAGGCGAGTTTTCGATTTGTCAGTACGCGG 600
Db 155 AAGTCGCGCGGAACTCCGCAATATCGAAGGCGCATTTTTCGATTTGTCAGTACGCGG 96
QY 601 GAAAGTACGCGGATCGCGGCGGACAGCGGCGCATTTTGGTGGTCCGCTTCGCGG 660
Db 95 GAAAGTTCGCGGATCGCGGCGGACAGCGGCGGATTTTGGTGGTCCGCTTCGCGG 36
QY 661 ATAAAGCGTCTTTTACGCGTCAAGCGTGGTGTGCAT 695
Db 35 ATAAAGCGTCTTTTACGCGTCAAGCGGCGGCAACAT 1

RESULT 13
AX002494/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
Direct Submission
Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1..13001
/organism="Neisseria meningitidis MC58"
/mol_type="genomic DNA"
/strains="MC58"
/db_xref="taxon:122586"
/notes="serogroup: B"

```



```

5892 GGGTCGGGACGGCTCTGCGTGTGGTCAATATTTGCGCGATATGCGCCGCGAATCCAAA 58933
|||||
541 AATGCGCGCGGAACTCGACGATATCGAAGGAGCAGCTTTTTCATTTGGTCAAGTACGGCG 600
|||||
5832 AAGTCGGCGCGGAACTCCGCAATATCGAAGGAGCAATATTTTTCATTTGGTCAAGTACGGCG 5773
|||||
601 GAAAGTACCGCGCGCATCGCGGCGGACAAAGCGGACGCGGATTTTGGTGGTCCGCTTCGG 660
|||||
5772 GAAAGTTCGGCGCATCGCGGCGGACAAAGCGGACGCGGATTTTGGTGGTCCGCTTCGG 5713
|||||
661 AATAAGGCTGCTTTTTCAGCGGTCTGAGTGTGCTGCAT 695
|||||
5712 ATAACGGTATTTTTCAGCAAGGCGGAGGACACAT 5678
|||||

RESULT 14
AX044032/c 349980 bp DNA linear PAT 24-NOV-2000
LOCUS
DEFINITION Sequence 111 from Patent WO0066791.
ACCESSION AX044032
VERSION AX044032.1 GI:11342916
KEYWORDS
SOURCE
ORGANISM
Neisseria meningitidis
Bacteria; Proteobacteria; Betaproteobacteria, Neisseriales;
Neisseriaceae; Neisseria.

REFERENCE
1
AUTHORS Pizza, M., Hickey, B., Peterson, J., Tettelin, H., Venter, J.C.,
Maignani, V., Galeotti, C., Mora, M., Ratti, G., Scarfelli, M.,
Scarlato, V., Rappuoli, R., Frazer, C.M. and Grandi, G.
TITLE Neisseria genomic sequences and methods of their use
JOURNAL Patent: WO 0066791-A 111 09-NOV-2000;
CHIRON CORPORATION (US); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
source
Location/Qualifiers
1. 349980
/organism="Neisseria meningitidis"
/mol_type="genomic DNA"
/db_xref="taxon:487"
/notes="sequence too long, cut in 8 pieces.-seq 1: 1 to
349980 349980 bases-seq 108: 300001 to 649980 349980
bases-seq 109: 600001 to 949980 349980 bases-seq 110:
900001 to 1249980 349980 bases-seq 111: 1200001 to 1549980
349980 bases-seq 112: 1500001 to 1849980 349980 bases-seq
113: 1800001 to 2149980 349980 bases-seq 114: 2100001 to
2272325 172325 bases"
BASE COUNT 86771 a 92803 c 86340 g 84066 t
ORIGIN

Query Match 21.9%; Score 639; DB 6; Length 349980;
Best Local Similarity 95.0%; Pred. No. 9.5e-124;
Matches 660; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 1 GAGTTTGGTTCGCGAAGCGAAGTGTGATGCTTGAGCGGACACCTGTCCGCGAAGCGG 60
DB 290504 GAGTTTGGTTCGCGAAGCGAAGTGTGATGCTTGAGCGGACACCTGTCCGCGAAGCGG 290445
QY 61 CTGACGCGCCCTTTTTCGCCCATCGACATCGTAACAAATCGGTTTGGTGGCAAGCTTTTC 120
DB 290444 CTGACGCGCCCTGTGCCCCATCGACATCGTAACAAATCGGTTTGGTGGCAAGCTTTTC 290385
QY 121 GCTTTGAGGTGCGAAGCGAAGTGTGATGCTTGAGCGGACACCTGTCCGCGTTCGCGCATCGCGCA 180
DB 290384 GCTTTGAGGTGCGGGAAGCGAAGTGTGATGCTTGAGCGGACACCTGTCCGCGTTCGCGCATCGCGCA 290325
QY 181 ATTTTCAGATGTCGCGCGCGAGTCTCTCATCTGTTTCAGACGCGCATACGATTTCTTCT 240
DB 290324 ATTTTCAGATGTCGCGCGCGAGTCTCTCATCTGTTTCAGACGCGCATACGATTTCTTCT 290265
QY 241 TCGCGCGCGGTGCGGTGAACACTCATGATTCAGACGAGCGCGCGATGCGGTTTTTTTGA 300
DB 290264 TCGCGCGCGGTGCGGTGAACACTCATGATTCAGACGAGCGCGCGATGCGGTTTTTTTGA 290205

```

```

QY 301 GCATGCGCCACGCGCGCGCGGACGCGGTTTTCGCGGAAAAAAGCTCGATATCGATAATG 360
DB 290204 GCATTTGCGCACGCGCGCACCGGACGCGGTTTTCGCGGAAAAAAGCTCGATATCGATAATG 290145
QY 361 TCGGCGCAGCGCGCTTTTCAATCAGCGAGTTCGAGCAGTTTCAAAATATATATTCGTCGGAACAC 420
DB 290144 TCGGCGCAGCGCGCTTTTCAATCAGCGCGCTTCGAGCAGTTTCAAAATATATATTCGTCGGAACAC 290085
QY 421 GGGAAACGAGCGCGCTTTCGCCATGCGCTGTGAACGTAACAGCAGCGGTTTTCGCGGCGACG 480
DB 290084 GGGAAACGAGCGCGCTTTCGCCATGCGCTGTGAACGTAACAGCAGCGGTTTTCGCGGCGACG 290025
QY 481 CGCTCGCGGACGCGTCTGCGTGTGCGCAATATTCGCGGATGCTGCGCGCAATTCGCGCAATTCGCGCA 540
DB 290024 CGCTCGCGGACGCGTCTGCGTGTGCGCAATATTCGCGGATGCTGCGCGCAATTCGCGCAATTCGCGCA 289965
QY 541 AAATCGCGCGGAACTTCGACGATATCGAAGGCGAGGTTTTCGATTTGGTCAAGTACGCGG 600
DB 289964 AAGTCGCGCGGAACTTCGCAATATCGAAGGCGATATTTTTCGATTTGGTCAAGTACGCGG 289905
QY 601 GAAAGTACGCGCGCATCGCGGCGGACGCGGCGGATTTTGGTGGTCCGCTTCG 660
DB 289904 GAAAGTTCGCGCGCATCGCGGCGGACGCGGCGGATTTTGGTGGTCCGCTTCG 289845
QY 661 ATAACGCGTGTTTTTCAGCGGTCTGCGTGTGTCAT 695
DB 289844 ATAACGCGTGTTTTTCAGCAAGGCGGAGGACACAT 289810

RESULT 15
AX001946
LOCUS
DEFINITION Deinococcus radiodurans R1 section 83 of 229 of the complete
chromosome 1
ACCESSION AE001946
VERSION AE001946.1 GI:6458655
KEYWORDS
SOURCE
ORGANISM
Deinococcus radiodurans
Deinococcus radiodurans
Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
Deinococcaceae; Deinococcus.
REFERENCE
1 (bases 1 to 10719)
AUTHORS White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,
Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,
Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,
Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C.,
Makarova, K.S., Aravind, L., Daly, M.J., Fraser, C.M. et al.
TITLE Genome sequence of the radioresistant bacterium Deinococcus
radiodurans R1
JOURNAL Science 286 (5444), 1571-1577 (1999)
MEDLINE 20036896
PUBMED 10567266
REFERENCE
2 (bases 1 to 10719)
AUTHORS White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,
Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,
Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,
Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C.,
Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W.,
Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S.,
Smith, H.O., Venter, J.C. and Fraser, C.M.
TITLE Direct Submission
JOURNAL Submitted (08-NOV-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
Location/Qualifiers
1. 10719
/organism="Deinococcus radiodurans"
/mol_type="genomic DNA"
/strains="R1"
/db_xref="taxon:1299"
/chromosome="1"
72. 1034
/gene="DR0932"
72. 1034
CDS

```



```
/gene="DR0932"
/Note="similar to SP:P19641 GB:X68873 PID:388220
PID:606125 GB:U00096 percent identity: 57.98; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="polyprenyl synthase"
/protein_id="AAFI0509.1"
/db_xref="GI:6458656"
/translacion="MPDAAFTRMEVLSRVSVEFTAIIGEDLVTAGGKVRPLVTLIA
AONLQAGDPAATLGHSHDELARIUSVCLLSHSLHDDLLDDSKRGQTAERRRC
NIVVWMSGDFVALRLLGELSGMPCSPALTRAFQQAASVCEGEVLQFOVAYAYSFE
NYFQVHKTKALLKELAAQAPMLLGGADDSARDALSTFGRYGFAPQDDLLDLGGE
EAQKPGKYGGLRBEKATYVPLCLGSPHDAEVAIRLERRAGEGDDVARMALAQER
TDDCTREIRRRIRRAIAALDALPSPPARTAALAEELARNT"
/1109..3043
/gene="DR0933"
/1109..3043
/gene="DR0933"
/Note="similar to PID:2094767 percent identity: 60.15;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="alpha-amyase"
/protein_id="AAFI0510.1"
/db_xref="GI:6458657"
/translacion="MLTPDLARLAFDDDDAETFLRLRYGPELADNLRAVYGN
HADALIGELLYVMHAYHARPADLKRDEALLRDLMDLQPEWGYVAYVDRFAGTLR
GVERLEYGLGVYLLHLLPLLRPDGENDGAVQVRSVRPDLGTYIDLSALARE
LRGRLSVLVDLVAHVAEHEWAVRATAGEAAYRDYFHIPDRTPDAYVEYFELF
PDFAGNLTNGEAGGWYTFNRSQMDVNGNPAVFRVEDLILTLANRGVEYFRLD
AIAFLKSLGTDCONQPEVRLTHALRAATRIAPAVAKAEIAPCGDLHYLGSRD
HIGRVSDVAYNSLVQVWSLSASRDTRLLFALAAFPFKPTNTNGVYVRCHDDIGW
AIADSDAAYNSLGAHGFPLSDFYSGFPGSFARGLPQHPQTGGRRISGTAAASLA
GLDIALETGDEWELVDAQADATGQVTPAGRNFAGLRHLLAVRRTPHLLHASTES
ADNRVHGFQMDVNLPAQADATGQVTPAGRNFAGLRHLLAVRRTPHLLHASTES
RPLPSPDCVLLLRHEHTGVLLQVNFSEHHTPTPTPLOEQLGAVHDLGSGSQPH
LGGPDLALPEPYALMLVAG"
/3065..3838
/gene="DR0934"
/3065..3838
/gene="DR0934"
/Note="similar to GB:AE000657 percent identity: 72.64;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, MerR family"
/protein_id="AAFI0511.1"
/db_xref="GI:6458658"
/translacion="MLVVCSTSLMGSDAKHRPVYVISAAELVDMHPOTLRLYERKGL
IRPSSGKTRLYSRDIEHLREIRLTOELGVNKGVEVMRLQHQDLDLQGFEEAE
TERJEDREQARPALPAPADPPDRPRPVYVISAELVDMHPOTLRLYERKOLIH
PGRSGKTRLYSRDIEHLREIRLTOELGVNKGVEVMRLQHQDLDLQGFEEAE
RQDDLSRMTWTWRLPAPSEAGPEEDAAADAGEDQN"
/3816..4907
/gene="DR0935"
/3816..4907
/gene="DR0935"
/Note="similar to GB:D10483 SP:P05637 GB:X04711 PID:216474
PID:40919 percent identity: 55.48; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="phosphatase, putative"
/protein_id="AAFI0512.1"
/db_xref="GI:6458659"
/translacion="MPEKTRTEHQBSLITFLRGVLLREPRFIVSDALSRSAFFVL
CEVFILNLAFLSPVPARPLWGVGDVGHGAKDKRALLRGLDGLDGLVGGHHL
VFLGDYVDRGDLGVRLVLSLEQALAGGEVFTALIGNEHMLAALFREQDPRD
RFGFDYWRNSGQGLTDLERLAPGLAWRLARPLAHAGQWLLHADSPPFYLGNDP
QEVNVRVALLISPDGPWAAAFANFVDRLXFDAGDGTAAQLRIETVGGRLVHGT
PTLLLQSGGDLPSGTPVLYADGICLAVDSGMAYFEDAGFVARIGELAPGGRABR
VLAELVTLPTPTFLDLSITATLRPSGEG"
4977..6017
/gene="DR0936"
4977..6017
/gene="DR0936"
/Note="similar to GB:AL009126 percent identity: 61.90;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAFI0513.1"
/db_xref="GI:6458660"
/translacion="MPLLLDHPDHYPLEGRKOLLEWVYVQGVTFDAMTNLPAGL
RALESYQNFPRFDIETVSHDGSVKILFTLNDGRQMEAVYMYLDRKTCVSTMGV
CARCACPATKMGSGRNLPGBIVGVQVAVAGGEGSPREIRNLVPMWGEALINTE
NSWAPARVLLHPDGLMGSKRVTLSVGIAGIRRLAEEDDLGKIAISLHAPDEETR
QQIIPCAANSIIBEINAAARDYQAVTGRVTMYTLMRGVNDHLWQELLALDGLV
SHVNLIPMNDGSDFFVSSSEEQIQAIFYDALQDRGVDSVRSRSGKDAAGACGQLALR
RPGARTGEVGA"
/6345..8135
/gene="DR0937"
/6345..8135
/gene="DR0937"
/Note="similar to PID:1652192 percent identity: 41.71;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="tetratricopeptide repeat family protein"
/protein_id="AAFI0514.1"
/db_xref="GI:6458661"
/translacion="MLVSPPTPRFLAARLLIGAWLACAGAAQOTLDTSAAGVQNT
LNQIQGVTVGSLSVQFQDACKPSCKQGPAGAPATSGQAQAQNAQAPAP
SOVAPATPTTPTTPEAGEAPATRVPTVPLTAEQQLLOAQQAQFAAGRYPOAR
NEFALLVRYNNPEPHFGGLGALYQLGLRGATFEFGQSQMOPAPQREYEPYLVIA
TREGRYDARLGEALTLAQSPAARQVLLRLALATEQGRKDYVALATYADLRAA
DPENVEYIFRAQALYQGVADALPVYVLERKPSLSQAALLLADLIYVAGQPDRA
VRELSAAARVLTGTRATLLRQSLLLAQNLRGALNASEARQEDRLPAALIRE
AELIAQGRQPAVAVRDALATPKSQAALYALYEGNOYPEARQAQQAQKLT
PDASTRAALYVOGVAVYRQGYTQARTALNSVLAVPDADSLMLGLSYVALKDYAS
APVLSVRLEPTEPAARONLAALATARYAEALLRGLVTDENVNAECWLLGLS
CRAQRPDDARQSPKTAAGLSARAKDALX"
/8132..9319
/gene="DR0938"
/8132..9319
/gene="DR0938"
/Note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAFI0516.1"
/db_xref="GI:6458663"
/translacion="MTRARTPSGPRSPRPRPAGRPWPPELVATLVLAGVGLGS
LLGERRAPABPGAPOTTISIYVSGAVVADGSPQTAAPQAQOTTAGKIPAPAPAP
APKIPPVLPPEPRMPAPPTPPRVERTTTAASTPTQPVTOAPATPTPTPAAQA
PATQPPATPAPBAPAPAPATPTTPEPAAPAGTPAVAPASQRTPLRSYRV
MLGTFGEAALRATAGVSAIGTIVYAIIDGNQFVAVQGVFADPESQQAADLRAY
ARALYPPRGCTLRNPAPAAAAPATPEPAPPTTPTTAPAPAGVYVQVGA
FNNVEGAQRFEQLRAQGFSPVNAPEYTKVTLVLLGTGSLTSTGRDLDAAGLDF
RLR"
/9316..10020
/gene="DR0939"
/9316..10020
/gene="DR0939"
/Note="similar to GB:AL009126 percent identity: 51.93;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAFI0515.1"
/db_xref="GI:6458662"
/translacion="MSRVADIPATVGRVYTVIRVBELEAQNVLRSASSELARRAGV
TPQVQRKOLYFGRGTGRTGTYVAVLRRELLRGLDQTNVNVIGMGRIGHAIY
PGASDYQFQVGLFDVAPDVVGVREVRGLTIQHSQGLGFPVAVSAGTPQVDMGLTVP
AETHAQAQAQALVAAGVGLNFAFVVLQTDQLHLPEAFAAGRREVTVENVDFLAGMK
```

1862	QY	AATGGGCAAGCTCGGAAACCTGCGCAGCGCACGCGCTCATCTCGCGGTTCAAATGC	1922
1960	DB	GCTGGGCACGACTCTGACAGACGCTTCAAGGCGGAGGCGAATGTGTGCGCGCGG	2019
1922	QY	CGTTATGCTGATTTGCCGCGCGCGCTGTCTTTCAAATCCGAAGCCATGCTCCACCCCGA	1981
2020	DB	GCGACCCGCGATCGTGTGCGCGCGCTTCAAGGCGGAGGCGAATGTGTGCGCGCGG	2079
1982	QY	CGAAGTCTGCAATACATCGGCGAGAGCAATGCCAA	2026
2080	DB	CGACTGATTCATTACCTCGGCTCGCGGACCAACATGGGCGGGTCTCGGACATGGCCTTA	2139
2027	QY	CAACCCCTGCAAATATGGCAATGTTGTGTGGAAACACCCCTTGCACGCGCGAAGTCAA	2086
2140	DB	CCAACACTCGCTGATGTGTGACGCTCTGGAGCAGCCTCGCCTCGCGTGACACGCGGCTGT	2199
2087	QY	CCATCAGGCGCTBA CCTACCGGCA CAA CCGTGCAGAGCATCCGCTGGGTCAACTACGT	2146
2200	DB	CACGCGCGCGCTTGGCGGCGTTTCCCGCCAAAGCCGACGAACGACTTGGGGCGTCTACGT	2259
2147	QY	CGCGAGCCACGACGACATCGGCTGTGACGTTTSCGATGAAGACGCGGCATATCTGGGCAT	2206
2260	DB	CGCGTGTCAOCAGCGACATCGGCTGGCGGATCCCGAOGAGGACGCGCGCGCTGGGCT	2319
2207	QY	AAGCGCTACGACACCGCCAACTTCTCAACCGCTTCTTCTGTCAACCGTTTTCGACGCGCAG	2266
2320	DB	CAGCGGACGGCGGACCGCCACTTCTTTGCGACTTTTACAGCGCGGAGTTTCCGCGCTC	2379
2267	QY	CTTCGCTGTGTGGGTACCGTTTCCATACAAACCAGACGACAGGCGACTGCGGTGTAGTGG	2326
2380	DB	CTTTGCGGGGACTGTGTTTTTCAGCACCCACCCACAGACCGGGGACCGCGCATTTTCGGG	2439
2327	QY	TACAGCGCGGCAATTGGTTCGGCTTG	2362
2440	DB	CACGCGGCGAGCCTTTCGCGGCTGCACTCGCGCTGGAAACGGGAGACGCGAGCGGGT	2499
2363	QY	TCCCCACGCGGTGACCGCATCAAACCTTGTACAGCAATGCTTTAGTACCGCGGCTCT	2422
2500	DB	GAAACGACGCGCTTGGCGGCTGCTGCTCTCTGACGCGGTCTGCTCGGCTTGGCGGGT	2559
2423	QY	GCGCTGTATTTACTTAGGCGACGAAGTGGGTACGCTCAATGACGAGCACTGTTCGACAGA	2482
2560	DB	GCGCTGTGTATCATGGGCGAGCACTCGCGTGTCTTAAACGACACTGACTTTTGGGCGCT	2619
2483	QY	CAGCAATAGACGCGACGACGCGGTTTGGCGCACCGTCCGCGCTACACGAAGCCCTGTA	2542
2620	DB	GCCGCGCCACGCGCGGCAACCGTGGGTTCAACCGGCGCGAGATGCACTGGAGCTGGT	2679
2543	QY	CGCGCAACCAAG	2584
2680	DB	CGCTCTGCGAGGCGGACGCGCGACCGGGCAGCGGTCACGCCCGCGGGCGAATGTT	2739
2585	QY	TACGGCTTTGGCGCATATGATTTGCCGCTTCGCCAAAGCAATCCGCGCTTCGACGGCGCA	2643
2740	DB	CSCCGGCTTGCAACACTGCTGCGGTGTGCGCGCGCACCCCGCACTCTGACGCCACGA	2798

Search completed: November 8, 2003, 16:01:00  
Job time : 10364 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 11:55:01 ; Search time 728 Seconds  
(without alignments)  
10805.168 Million cell updates/sec

Title: US-09-843-007A-1

Perfect score: 2914

Sequence: 1 GAGTTTCGCTCCGACCC.....CGTTTCAGCGCATTTGCG 2914

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq 19Jun03:\*

- 1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*
- 25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2914	100.0	2914	21	AAA11732
2	2914	100.0	4173	21	AAA07380
3	2641	90.6	2883	17	AAT09860
4	2641	90.6	2883	17	AA111179
5	1921.4	65.9	1939	22	AAF61709
6	1921.4	65.9	6878	22	AAF61711
7	1891.4	64.9	1310	22	AAF61710
8	1891.4	64.9	6851	22	AAF61712

C	9	655	22.5	795	24	AAS97236
C	10	639	21.9	765	25	ABX09900
C	11	639	21.9	48275	21	AAA81501
C	12	639	21.9	349980	21	AAF21610
C	13	639	21.9	837096	21	AAA81489
C	14	637.4	21.9	762	25	ABZ40085
C	15	297.2	10.2	29559	23	AAS59546
C	16	169	5.8	759	23	AAS52439
C	17	161.4	5.5	882	23	AAS56202
C	18	153.2	5.3	1163020	24	ABQ67197
C	19	153.2	5.3	3011208	24	ABQ69245
C	20	151.6	5.2	799	12	AAQ15228
C	21	150	5.1	2944528	24	ABA03041
C	22	147.4	5.1	1340	24	ABQ70237
C	23	142.6	4.9	801	23	AAS31122
C	24	142.6	4.9	6415	20	AAH13212
C	25	142.6	4.9	6415	24	ABX99007
C	26	133	4.6	756	24	ABQ68513
C	27	120.4	4.1	3303	23	AAS54127
C	28	118	4.0	483	24	ABK73108
C	29	112.2	3.9	1126	23	AAS29731
C	30	111.8	3.8	1794	22	AAH67492
C	31	111.8	3.8	1981	24	ABX85343
C	32	111.8	3.8	349980	22	AAH68531
C	33	111.8	3.8	349980	22	AAH68532
C	34	104.6	3.6	1102	22	AAF71764
C	35	104.4	3.6	1704	17	AAT11413
C	36	99.4	3.4	4403765	22	AAI99683
C	37	99.4	3.4	4411529	22	AAI99682
C	38	91.8	3.2	1113	23	AAS99702
C	39	85.8	2.9	22934	23	AAS99613
C	40	82.2	2.8	293	24	ABN21665
C	41	76.2	2.6	349980	24	ABQ81845
C	42	73.6	2.5	3600	17	AAT13227
C	43	65	2.2	1689	24	ABK73366
C	44	61.6	2.1	2238	23	AAS27660
C	45	59.8	2.1	1782	16	AAT03251

## ALIGNMENTS

### RESULT 1

AAA11732

ID AAA11732 standard; DNA; 2914 BP.

XX AAA11732;

XX 21-JUL-2000 (first entry)

XX N. polysacchara amylosucrase DNA.

XX Alpha-1,6-branched alpha-1,4-glucan; branching enzyme; binder; carrier;  
 KW transgenic plant; flavor; perfume; packaging material; papermaking;  
 KW ultra-violet light adsorber; starch; textile; wetting agent;  
 KW amylosucrase; ds.

XX Neisseria polysacchara.

XX Key Location/Qualifiers

XX CDS 957..2867

XX /\*tag= a

XX /product= "amylosucrase"

XX WO2000022140-A1.

XX 20-APR-2000.

XX 08-OCT-1999; 99WO-EP07562.

XX 09-OCT-1998; 98DE-1046635.

XX 27-MAY-1999; 99DE-1024342.

XX

PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.  
XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
PI Buettcher V, Quanz M;  
XX WPI; 2000-317992/27.  
DR P-PSDB; AAW90979.  
XX  
PT New nucleic acid encoding a branching enzyme, useful for in vitro  
PT synthesis of branched glucans and to prepare transgenic plants  
PT producing modified starch -  
XX  
PS Disclosure; Page 99-102; 115pp; German.  
XX  
CC This invention describes a novel nucleic acid (I) isolated from  
CC Neisseria which encodes a branching enzyme (II). (I) is used for  
CC recombinant production of (II) subsequently used in the in vitro  
CC production of alpha-1,6-branched alpha-1,4-glucans. It is also used to  
CC prepare transgenic plants that produce starches with modified properties.  
CC (III) are used as binders for tablets, carriers for pharmaceuticals,  
CC flavors and perfumes and powdered additives, packaging materials,  
CC ultra-violet light adsorbers in sunscreens and also for any of the usual  
CC applications of starch in foods, papermaking, as textile size, in soil  
CC stabilization, as wetting agent for agricultural chemicals, as polymer  
CC additives etc. Fragments of (I) are useful as PCR primers and antisense  
CC molecules or ribozymes for inhibiting expression of (I), and the  
CC regulatory region of (II) can be used to control expression of  
CC heterologous sequences in host cells. (I) provides an inexpensive method  
CC for producing alpha-1,6-branched alpha-1,4-glucans (III), producing  
CC products that can be tailored for particular applications, particularly  
CC by controlling the degree of branching. Starch from transgenic plants  
CC has increased gel strength; reduced phosphate content; reduced peak  
CC viscosity; lower pasting temperature and granule size and/or altered  
CC sidechain distribution. This sequence encodes an amylase isolated  
CC from Neisseria polysacchara which is described in the method of the  
CC invention.  
XX  
SQ Sequence 2914 BP; 662 A; 868 C; 773 G; 611 T; 0 other;  
Query Match 100.0%; Score 2914; DB 21; Length 2914;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2914; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GAGTTTGGTTCCTCCGAAACGGAACGATGATGCTTGGAGCGGAACACCTGTCCGCGCAAGGGCGG 60  
DB 1 GAGTTTGGTTCCTCCGAAACGGAACGATGATGCTTGGAGCGGAACACCTGTCCGCGCAAGGGCGG 60  
QY 61 CTGACCGCCCTTTTGGCCCATCGACATCGTAACAAATCGTTTGGTGGCAAGCTCTTTC 120  
DB 61 CTGACCGCCCTTTTGGCCCATCGACATCGTAACAAATCGTTTGGTGGCAAGCTCTTTC 120  
QY 121 GCTTTGAGCGTGGCAGAAAGCAAGTCAGACGTCTTTCGCGTTTGGCGCATCACCGCA 180  
DB 121 GCTTTGAGCGTGGCAGAAAGCAAGTCAGACGTCTTTCGCGTTTGGCGCATCACCGCA 180  
QY 181 ATTTTCAGATGTCGCGCGGCGAGTCTCCATCTGTTTCAGACGGCATACGATTTCTTC 240  
DB 181 ATTTTCAGATGTCGCGCGGCGAGTCTCCATCTGTTTCAGACGGCATACGATTTCTTC 240  
QY 241 TCGCGGCGGTGGGTGAACCTCATGATTCGACAGCAGCGCGCGATCGGTTTTCGA 300  
DB 241 TCGCGGCGGTGGGTGAACCTCATGATTCGACAGCAGCGCGCGATCGGTTTTCGA 300  
QY 301 GCATGGCCACCGCGCGCGCGGAGTTCGCGGAAAGCTCGATATCGATAATG 360  
DB 301 GCATGGCCACCGCGCGCGCGGAGTTCGCGGAAAGCTCGATATCGATAATG 360  
QY 361 TCGCGGCGGCGGTTCATCAGAGTCGAGTCAGATCAAAATTAATCGTCGGAAC 420  
DB 361 TCGCGGCGGCGGTTCATCAGAGTCGAGTCAGATCAAAATTAATCGTCGGAAC 420  
QY 421 GGGAAACGAGCGGCTTCGCCATCGCTCGAAACGTAACAGCAGCGGCTTGTCCGCGAC 480  
DB 421 GGGAAACGAGCGGCTTCGCCATCGCTCGAAACGTAACAGCAGCGGCTTGTCCGCGAC 480  
QY 481 GCGTCGCGGAGCGGTCTCGGTGTGGCGCAATCTTCGCGGATGTCGCGGCAATTCACAA 540  
DB 481 GCGTCGCGGAGCGGTCTCGGTGTGGCGCAATCTTCGCGGATGTCGCGGCAATTCACAA 540  
QY 541 AAATCGCGCGGAACTCGACGATATCGAAGGCGAGGTTTTTGGATTGGTCAAGTACGCG 600  
DB 541 AAATCGCGCGGAACTCGACGATATCGAAGGCGAGGTTTTTGGATTGGTCAAGTACGCG 600  
QY 601 GAAATGACGGCGGATCGCGGCGACAGCGGCGATTTTGGTGGCGTCCGCTTCG 660  
DB 601 GAAATGACGGCGGATCGCGGCGACAGCGGCGATTTTGGTGGCGTCCGCTTCG 660  
QY 661 ATAAACGCTGTTTTGACGGTCTGCTGTGTCATGCGGTTGTCGGGTGAAAGGAAC 720  
DB 661 ATAAACGCTGTTTTGACGGTCTGCTGTGTCATGCGGTTGTCGGGTGAAAGGAAC 720  
QY 721 GGTAAAGCAATATAGCAAGGCAACGCGATGTTTTCAGACGGCATTTCTGTGGCG 780  
DB 721 GGTAAAGCAATATAGCAAGGCAACGCGATGTTTTCAGACGGCATTTCTGTGGCG 780  
QY 781 CGGCTTGATATGAATCAAGCAGCATCGCATATCGAATCGAGACTTTGGCAAGCCCTG 840  
DB 781 CGGCTTGATATGAATCAAGCAGCATCGCATATCGAATCGAGACTTTGGCAAGCCCTG 840  
QY 841 TCTTTTCTAGTCAGTCGCGAGTCTTTCAGTATGATTCGACGACACGCCCTTACAGCAT 900  
DB 841 TCTTTTCTAGTCAGTCGCGAGTCTTTCAGTATGATTCGACGACACGCCCTTACAGCAT 900  
QY 901 TTGAGGATATCGCGCGGAGACCGCGGTTCGGAACCTTCAGAAATCGAGCAGCATCATGT 960  
DB 901 TTGAGGATATCGCGCGGAGACCGCGGTTCGGAACCTTCAGAAATCGAGCAGCATCATGT 960  
QY 961 TGACCCCGCAGCAGCAAGTCGTTGATTTTACAGTACCTCAAAACAGCATCTTGGACA 1020  
DB 961 TGACCCCGCAGCAGCAAGTCGTTGATTTTACAGTACCTCAAAACAGCATCTTGGACA 1020  
QY 1021 TCTACAGCCCGAAACAGCGCGCGCATCGAATAATCGAAGACTGGCGGCGATTTTCGC 1080  
DB 1021 TCTACAGCCCGAAACAGCGCGCGCATCGAATAATCGAAGACTGGCGGCGATTTTCGC 1080  
QY 1081 GCGCATGATGATGCGATTTTCCCAAACTGATGATGATGATGATGATGATGATGATGATG 1140  
DB 1081 GCGCATGATGATGCGATTTTCCCAAACTGATGATGATGATGATGATGATGATGATGATG 1140  
QY 1141 ACGAGCCCTGCTCTCTATGCTGGAATGCTGCTGGCGAGGATGCGAAGCTATTTCCC 1200  
DB 1141 ACGAGCCCTGCTCTCTATGCTGGAATGCTGCTGGCGAGGATGCGAAGCTATTTCCC 1200  
QY 1201 AACGCAACTCTCTCTTAAAGATATCGATATCGCGCGCGAAGAAACAAACCCGATTTG 1260  
DB 1201 AACGCAACTCTCTTAAAGATATCGATATCGCGCGCGAAGAAACAAACCCGATTTG 1260  
QY 1261 TGTCCAAACAAAGTCGCGCGGCTGCTAGCTGATTTGCTTTGCGCGGATTTGAAGG 1320  
DB 1261 TGTCCAAACAAAGTCGCGCGGCTGCTAGCTGATTTGCTTTGCGCGGATTTGAAGG 1320  
QY 1321 GCTTGAAGATATAAATCTCTTATTTTCAAGAGCTTGGTTTTCATCTATCTGACCTGATGC 1380  
DB 1321 GCTTGAAGATATAAATCTCTTATTTTCAAGAGCTTGGTTTTCATCTATCTGACCTGATGC 1380  
QY 1381 GCTGTTTAAATGCTGAAAGCAAGCGCGGCTATGCGTTCAGCTACCGG 1440  
DB 1381 GCTGTTTAAATGCTGAAAGCAAGCGCGGCTATGCGTTCAGCTACCGG 1440  
QY 1441 ATGTCAATCCGCGCATCGGCGCAATAGCGGCTTTCGCGGAGTCAFTGCTGCGTGCACG 1500  
DB 1441 ATGTCAATCCGCGCATCGGCGCAATAGCGGCTTTCGCGGAGTCAFTGCTGCGTGCACG 1500  
QY 1501 AAGCGCGCATTTCCGCGCTGCTGATTTTATCTTCAACACACCTTCCCAACAGCAAT 1560  
DB 1501 AAGCGCGCATTTCCGCGCTGCTGATTTTATCTTCAACACACCTTCCCAACAGCAAT 1560

QY	1561	GGGCGCAACGTTGCGCGCGCGGACACCGGCTTTTCGA	CAATTTCTATA	TATATTTCCCG	1620
DB	1561	GGGCGCAACGTTGCGCGCGCGGACCGGCTTTTCGA	CAATTTCTATA	TATATTTCCCG	1620
QY	1621	ACCGCGGATGCGGACGAATACGACCGCACCTGCGGAAATCTTCCCGACGACCC	1680		
DB	1621	ACCGCGGATGCGGACGAATACGACCGCACCTGCGGAAATCTTCCCGACGACCC	1680		
QY	1681	CGGCGGCTTCTCGCAACTGGAACAGCGAGCGCTGGGTGTGGAGACCTTCAATTCCTTCC	1740		
DB	1681	CGGCGGCTTCTCGCAACTGGAAGAGGAGCGCTGGGTGTGGAGACCTTCAATTCCTTCC	1740		
QY	1741	AATGGGACTTTGAATTA	CAGCAACCGGTGGGTATTC	CGCGCAATGGGGGCGCAATCGTGT	1800
DB	1741	AATGGGACTTTGAATTA	CAGCAACCGGTGGGTATTC	CGCGCAATGGGGGCGCAATCGTGT	1800
QY	1801	TCCTTTGCCAACTTGGGGGTTTGACATCTCTCGGTATG	GTGCGTTGCGCTTTATTTGGAAC	1860	
DB	1801	TCCTTTGCCAACTTGGGGGTTTGACATCTCTCGGTATG	GTGCGTTGCGCTTTATTTGGAAC	1860	
QY	1861	AAATGGGGACAAGCTGGGAAACCTGCGCGAGCGCACGCGCTCATCGCGCGCTTCAATG	1920		
DB	1861	AAATGGGGACAAGCTGGGAAACCTGCGCGAGCGCACGCGCTCATCGCGCGCTTCAATG	1920		
QY	1921	CCGTTATGCGTATTTGCGCGCGCGCGCTGTTCTTCAAATCGAAGCCATCGTCCACCCCG	1980		
DB	1921	CCGTTATGCGTATTTGCGCGCGCGCGCTGTTCTTCAAATCGAAGCCATCGTCCACCCCG	1980		
QY	1981	ACCAAGTGTGCCAATACATCGGCGAGACGAATCGCAANTCGGTTACAA	CGCCCTGCAAA	2040	
DB	1981	ACCAAGTGTGCCAATACATCGGCGAGACGAATCGCAANTCGGTTACAA	CGCCCTGCAAA	2040	
QY	2041	TGGCAATTTGTTGGAAACACCTTGCCACCGCGGGAAGTCAACCTGCTCCATCAGCGCGCTGA	2100		
DB	2041	TGGCAATTTGTTGGAAACACCTTGCCACCGCGGGAAGTCAACCTGCTCCATCAGCGCGCTGA	2100		
QY	2101	CCTACCGGCACAACTGTCGCGAGCATACGCGCTGGGTCAACTACGTCGCGAGCACGACG	2160		
DB	2101	CCTACCGGCACAACTGTCGCGAGCATACGCGCTGGGTCAACTACGTCGCGAGCACGACG	2160		
QY	2161	ACATCGGCTGACGCTTTGCCGATGAAGACGCGGCATATCTGGGCATAAAGCGGTTACGAC	2220		
DB	2161	ACATCGGCTGACGCTTTGCCGATGAAGAGCGCGCATATCTGGGCATAAAGCGGTTACGAC	2220		
QY	2221	ACCGCCAAATTCCTCAAACCGCTTCTTGCTCAACCGTTTCGACGCGAGCTTCGCTCGTGGCG	2280		
DB	2221	ACCGCCAAATTCCTCAAACCGCTTCTTGCTCAACCGTTTCGACGCGAGCTTCGCTCGTGGCG	2280		
QY	2281	TACCGTTCCCAATACAA	CCCAAGCACAGCGCACTCGCGTGCAGTGTGTACAGCGCGCGGCAT	2340	
DB	2281	TACCGTTCCCAATACAA	CCCAAGCACAGCGCACTCGCGTGTGCAGTGTGTACAGCGCGCGGCAT	2340	
QY	2341	TGGTCGGCTTGGCGCAAGACGATCCCAACGCGCTTGA	CCGGATCAAA	CTCTGTGTACGCA	2400
DB	2341	TGGTCGGCTTGGCGCAAGACGATCCCAACGCGCTTGA	CCGGATCAAA	CTCTGTGTACGCA	2400
QY	2401	TTGCTTTTGGTACCGCGCGCTCTGCGCGTGTATTAC	TGAGCGACGAAGTGGGTACGCTCA	2460	
DB	2401	TTGCTTTTGGTACCGCGCGCTCTGCGCGTGTATTAC	TGAGCGACGAAGTGGGTACGCTCA	2460	
QY	2461	ATGACGACGACTGCTGCAAGACGCAATATAGACGCA	CGACGCGCTTGGGCGCACCGCTC	2520	
DB	2461	ATGACGACGACTGCTGCAAGACGCAATATAGACGCA	CGACGCGCTTGGGCGCACCGCTC	2520	
QY	2521	CGCGCTACAA	CGAAGCCCTGTACGCGCAACGCAACGATCCGTCCACCGCACGCGCGGCAAA	2580	
DB	2521	CGCGCTACAA	CGAAGCCCTGTACGCGCAACGCAACGATCCGTCCACCGCACGCGCGGCAAA	2580	
QY	2581	TCTATCAGGGCTTGGCGCATATGATTTGCGTCCGCGCA	AAAGCAATCCCGCGCTTCGACGCGCG	2640	
DB	2581	TCTATCAGGGCTTGGCGCATATGATTTGCGTCCGCGCA	AAAGCAATCCCGCGCTTCGACGCGCG	2640	

## RESULT 2

RESULI 2  
AAA07380

AAA00380  
ID AAA07380 standard: DNA; 4173 BP.

AC AAA07380:

XX  
100510557

DT 30-JUN-2000 (first entry)

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

DE *Neisseria polysaccharea* amylosucrase coding sequence.

XX

KW Amylosucrase; I

KW extracellular

XXV

OS *Neisseria polysaccharea*.

XX  
1234567890 31

PR - WO200014249-A1.  
YY

XX 0000-0000-0000

XX  
FD  
16-MAR-2000.

02-SBP-1998: 98WO-EP05573.

XX  
20-25-30-35-40-45-50-55-60-65-70-75-80-85-90-95-100-105-110-115-120-125-130-135-140-145-150-155-160-165-170-175-180-185-190-195-200-205-210-215-220-225-230-235-240-245-250-255-260-265-270-275-280-285-290-295-300-305-310-315-320-325-330-335-340-345-350-355-360-365-370-375-380-385-390-395-400-405-410-415-420-425-430-435-440-445-450-455-460-465-470-475-480-485-490-495-500-505-510-515-520-525-530-535-540-545-550-555-560-565-570-575-580-585-590-595-600-605-610-615-620-625-630-635-640-645-650-655-660-665-670-675-680-685-690-695-700-705-710-715-720-725-730-735-740-745-750-755-760-765-770-775-780-785-790-795-800-805-810-815-820-825-830-835-840-845-850-855-860-865-870-875-880-885-890-895-900-905-910-915-920-925-930-935-940-945-950-955-960-965-970-975-980-985-990-995-1000-1005-1010-1015-1020-1025-1030-1035-1040-1045-1050-1055-1060-1065-1070-1075-1080-1085-1090-1095-1100-1105-1110-1115-1120-1125-1130-1135-1140-1145-1150-1155-1160-1165-1170-1175-1180-1185-1190-1195-1200-1205-1210-1215-1220-1225-1230-1235-1240-1245-1250-1255-1260-1265-1270-1275-1280-1285-1290-1295-1300-1305-1310-1315-1320-1325-1330-1335-1340-1345-1350-1355-1360-1365-1370-1375-1380-1385-1390-1395-1400-1405-1410-1415-1420-1425-1430-1435-1440-1445-1450-1455-1460-1465-1470-1475-1480-1485-1490-1495-1500-1505-1510-1515-1520-1525-1530-1535-1540-1545-1550-1555-1560-1565-1570-1575-1580-1585-1590-1595-1600-1605-1610-1615-1620-1625-1630-1635-1640-1645-1650-1655-1660-1665-1670-1675-1680-1685-1690-1695-1700-1705-1710-1715-1720-1725-1730-1735-1740-1745-1750-1755-1760-1765-1770-1775-1780-1785-1790-1795-1800-1805-1810-1815-1820-1825-1830-1835-1840-1845-1850-1855-1860-1865-1870-1875-1880-1885-1890-1895-1900-1905-1910-1915-1920-1925-1930-1935-1940-1945-1950-1955-1960-1965-1970-1975-1980-1985-1990-1995-2000-2005-2010-2015-2020-2025-2030-2035-2040-2045-2050-2055-2060-2065-2070-2075-2080-2085-2090-2095-2100-2105-2110-2115-2120-2125-2130-2135-2140-2145-2150-2155-2160-2165-2170-2175-2180-2185-2190-2195-2200-2205-2210-2215-2220-2225-2230-2235-2240-2245-2250-2255-2260-2265-2270-2275-2280-2285-2290-2295-2300-2305-2310-2315-2320-2325-2330-2335-2340-2345-2350-2355-2360-2365-2370-2375-2380-2385-2390-2395-2400-2405-2410-2415-2420-2425-2430-2435-2440-2445-2450-2455-2460-2465-2470-2475-2480-2485-2490-2495-2500-2505-2510-2515-2520-2525-2530-2535-2540-2545-2550-2555-2560-2565-2570-2575-2580-2585-2590-2595-2600-2605-2610-2615-2620-2625-2630-2635-2640-2645-2650-2655-2660-2665-2670-2675-2680-2685-2690-2695-2700-2705-2710-2715-2720-2725-2730-2735-2740-2745-2750-2755-2760-2765-2770-2775-2780-2785-2790-2795-2800-2805-2810-2815-2820-2825-2830-2835-2840-2845-2850-2855-2860-2865-2870-2875-2880-2885-2890-2895-2900-2905-2910-2915-2920-2925-2930-2935-2940-2945-2950-2955-2960-2965-2970-2975-2980-2985-2990-2995-3000-3005-3010-3015-3020-3025-3030-3035-3040-3045-3050-3055-3060-3065-3070-3075-3080-3085-3090-3095-3100-3105-3110-3115-3120-3125-3130-3135-3140-3145-3150-3155-3160-3165-3170-3175-3180-3185-3190-3195-3200-3205-3210-3215-3220-3225-3230-3235-3240-3245-3250-3255-3260-3265-3270-3275-3280-3285-3290-3295-3300-3305-3310-3315-3320-3325-3330-3335-3340-3345-3350-3355-3360-3365-3370-3375-3380-3385-3390-3395-3400-3405-3410-3415-3420-3425-3430-3435-3440-3445-3450-3455-3460-3465-3470-3475-3480-3485-3490-3495-3500-3505-3510-3515-3520-3525-3530-3535-3540-3545-3550-3555-3560-3565-3570-3575-3580-3585-3590-3595-3600-3605-3610-3615-3620-3625-3630-3635-3640-3645-3650-3655-3660-3665-3670-3675-3680-3685-3690-3695-3700-3705-3710-3715-3720-3725-3730-3735-3740-3745-3750-3755-3760-3765-3770-3775-3780-3785-3790-3795-3800-3805-3810-3815-3820-3825-3830-3835-3840-3845-3850-3855-3860-3865-3870-3875-3880-3885-3890-3895-3900-3905-3910-3915-3920-3925-3930-3935-3940-3945-3950-3955-3960-3965-3970-3975-3980-3985-3990-3995-4000-4005-4010-4015-4020-4025-4030-4035-4040-4045-4050-4055-4060-4065-4070-4075-4080-4085-4090-4095-4100-4105-4110-4115-4120-4125-4130-4135-4140-4145-4150-4155-4160-4165-4170-4175-4180-4185-4190-4195-4200-4205-4210-4215-4220-4225-4230-4235-4240-4245-4250-4255-4260-4265-4270-4275-4280-4285-4290-4295-4300-4305-4310-4315-432

PR 02-SEP-1998; 98WO-EP05573.

XXXXXX

PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.




PI Quanz M, Provart N;

XX

DR WPI; 2000-256995/22.

DR P-PSDB; AAY81945.

XX  
nm

Various commercial polymers obtained from

Novel amylosucrase polynucleotides and polypeptides obtained from *Neisseria polysaccharae* used for production of alpha-1,4-glucans.

PT *Neisseria polysaccharea*, used for production of alpha-1,3 glucans,  
fructose and cyclodextrins -

Fructose and cyclodextrins -  
PI  
xx

PS Claim 1: Page 25-30; 36pp: English.

XX  
For CHAM 1, Page 23 of 30, Copy, margin.

CC This sequence encodes the *Neisseria polysa*

invention. The amylosucrose sequences are



```
Db 3175 ACATCGGCTGAGCGTTTCCCGAGTAAAGACGCGGCATATCTGGGCATTAAGCGGCTACGACC 3234
Qy 2221 ACCGCCAATTCCTCAACCGCTTCTTCGTCAACCGCTTTCGACGGCAGCTTCGCTCGTGGCG 2280
Db 3235 ACCGCCAATTCCTCAACCGCTTCTTCGTCAACCGCTTTCGACGGCAGCTTCGCTCGTGGCG 3294
Qy 2281 TACCGTTCCNATACAAACCAAGCAGGAGTCCGCTGTCAGTGTGTCAGCGCGGCGAT 2340
Db 3295 TACCGTTCCNATACAAACCAAGCAGGAGTCCGCTGTCAGTGTGTCAGCGCGGCGAT 3354
Qy 2341 TGGTCGGCTTGGCGCAAGACGATCCGCCAGCGCTTTCGACGCATCAAACTCTTTGTACAGCA 2400
Db 3355 TGGTCGGCTTGGCGCAAGACGATCCGCCAGCGCTTTCGACGCATCAAACTCTTTGTACAGCA 3414
Qy 2401 TTGCTTTGAGTACCGCGGCTTCGCGCTGATTTACTTAGGCGAGCAAGTGGTACGCTCA 2460
Db 3415 TTGCTTTGAGTACCGCGGCTTCGCGCTGATTTACTTAGGCGAGCAAGTGGTACGCTCA 3474
Qy 2461 ATGACGACGACTGTGCGCAAGCAGCAATTAAGAGCGACGACAGCGCTTTGGGCGCACCGTC 2520
Db 3475 ATGACGACGACTGTGCGCAAGCAGCAATTAAGAGCGACGACAGCGCTTTGGGCGCACCGTC 3534
Qy 2521 CGCGTACAAAGAACCCCTGTACGGCGAAACGCAACGATCCGTCGACCGAGCCGGGCAAA 2580
Db 3535 CGCGTACAAAGAACCCCTGTACGGCGAAACGCAACGATCCGTCGACCGAGCCGGGCAAA 3594
Qy 2581 TCTATCAGGCTTGGCGCATATGATTCGCTCCGCCAAAGCAATCCGCGCTTCGACGCG 2640
Db 3595 TCTATCAGGCTTGGCGCATATGATTCGCTCCGCCAAAGCAATCCGCGCTTCGACGCG 3654
Qy 2641 GCAGCTGGTTACATTCACACCAACCAAGCAGCATCATCGGCTACATTCGCGCAACATG 2700
Db 3655 GCAGCTGGTTACATTCACACCAACCAAGCAGCATCATCGGCTACATTCGCGCAACATG 3714
Qy 2701 CGCTTTGGCATTCGTAACCTCAGGATATTCGCAACCAACCGTTACCGGATACCCCTGC 2760
Db 3715 CGCTTTGGCATTCGTAACCTCAGGATATTCGCAACCAACCGTTACCGGATACCCCTGC 3774
Qy 2761 AAGCATGCCCTTCAAGGCGACGACCTCATCGCTGGCAAACTCTGACGCTGATCAGG 2820
Db 3775 AAGCATGCCCTTCAAGGCGACGACCTCATCGCTGGCAAACTCTGACGCTGATCAGG 3834
Qy 2821 ATTTGACGCTTACGCGCTTACGATGTCAGTCAATGCTGCTGAAATGCGCTGACGACGCTTCCA 2880
Db 3835 ATTTGACGCTTACGCGCTTACGATGTCAGTCAATGCTGCTGAAATGCGCTGACGACGCTTCCA 3894
Qy 2881 AATGCGCTGTAACCGTTTTCAGACGCGATTTCG 2914
Db 3895 AATGCGCTGTAACCGTTTTCAGACGCGATTTCG 3928
```

## RESULT 3

AAT09860

ID AAT09860 standard; DNA; 2883 BP.

XX AC AAT09860;

XX DT 08-APR-1996 (first entry)

XX DE Neisseria polysaccharea amylosucrase DNA sequence.

XX KW Amylosucrase; enzyme; amylose, linear 1,4-glucan; transgenic plant;

XX KW EC-2.4.1.4; crop improvement; ss.

XX OS Neisseria polysaccharea.

XX FH Key Location/Qualifiers

XX FT 5'UTR 1..938

XX FT /tag= a

XX FT sig\_peptide 939..986

XX FT /\*tag= b

XX FT 939..2780

XX FT CDS /\*tag= c

XX W09531553-A1.

XX 23-NOV-1995.

XX 18-MAY-1995; 95WO-EP01893.

XX 22-DEC-1994; 94DB-4447388.

XX 18-MAY-1994; 94DB-4417879.

XX (GENB-) INST GENBIOLOGISCHE FORSCHUNG.

XX Buettcher V, Kossmann J, Welsh T;

XX WPI; 1996-010938/01.

XX P-FSDB; AAR88386.

XX DNA encoding amylo:sucrase from *Neisseria polysaccharea* - for prodn. of linear 1,4-glucan(s), esp. amylose, from sucrose.

XX Claim 1; Page 39-42; 56pp; English.

XX This DNA sequence encodes an amylosucrase which allows the synthesis of linear alpha-1,4-glucans from the substrate sucrose by bacteria, fungi and plants, or in cell-free systems. This sequence may be expressed recombinantly.

XX Sequence 2883 BP; 661 A; 852 C; 758 G; 612 T; 0 other;

Query Match 90.6%; Score 2641; DB 17; Length 2883;

Best Local Similarity 97.8%; Pred. No. 0;

Matches 2852; Conservative 0; Mismatches 30; Indels 33; Gaps 16;

Qy 1 GAGTTTTCGCTTCCCGAACCGAAGCTGATGTTGAGCGGAAACCTGTTCGGCAAGCGG 60

Db 1 GAGTTTTCGCTTCCCGAACCGAAGCTGATGTTGAGCGGAAACCTGTTCGGCAAGCGG 60

Qy 61 CTGACCGGCCCCCTTTGCGCCCATCGACATCGTAACATCGTTCGTCGCAAGCTCTTTC 120

Db 61 -----ACCGCTTTTTCGCCCATCGACATCGTAACATCGTTCGTCGCAAGCTCTTTC 114

Qy 121 GCTTTGAGCGTGGCAGAAAGCAAGTTCAGCAGCTTCCTCCGCGCTTTCGGGCAACCGCA 180

Db 115 GCTTTGAGCGTGGCAGAAAGCAAGTTCAGCAGCTTCCTCCGCGCTTTCGGGCAACCGCA 173

Qy 181 ATTTTGACAGTTCGCGCGCGCATCTCCATCTGTTTTCAGACGGCATACGATTCTTCT 240

Db 174 ATTTTGACAGTTCGCGCGCGCATCTCCATCTGTTTTCAGACGGCATACGATTCTTCT 233

Qy 241 TCGCGCGCGCTGGCTGAACTCATGATTCAGACGGCGCGGCGATGCGCTTTTTTGA 300

Db 234 TCGCGCGCGCTGGCTGAACTCATGATTCAGACGGCGCGGCGATGCGCTTTTTTGA 290

Qy 301 GCATCGCGCCACGCGCGCGCGGCTTTCGCGGAAAGCTTCGATATCGATAATG 360

Db 291 GCATCGCGCCACGCGCGCGG-----AGCGTTTTCGCGGAAAGCTTCGATATCGATAATG 344

Qy 361 TCGCGCGCGCGCTTTCATCAGCGAGTTCGAGCTTCGAAATATATATCGTTCGGAACAC 420

Db 345 TCGCGCGCGCGCTTTCATCAGCGAGTTCGAGCTTCGAAATATATATCGTTCGGAACAC 404

Qy 421 GCGAACAGCGCGCTTCGCCATGCGCTCTCAACGTAACAGCAGCGGCTTTCGCGGACG 480

Db 405 GCGAACAGCGCGCTTCGCCATGCGCTCTCAACGTAACAGCAGCGGCTTTCGCGGACG 464

Qy 481 CGCTCGCGGACGCTTCGCTGTGGCGAATCTTCGCGGATGCTTCGCGCGCATTCGCAAA 540

Db 465 CGCTCGCGGACGCTTCGCTGTGGCGAATCTTCGCGGATGCTTCGCGCGCATTCGCAAA 524

Qy 541 AAATCGCGCGGAACTCGACGATATCGAAGGCGAGGTTTTGATTGCTCAAGTACGCGG 600

Db 525 AAATCGCGCGGAACTCGACGATATCGAAGGCGAGGTTTTGATTGCTCAAGTACGCGG 584





Db 2729 CAAGCGATGCCCTTCAGCGGCGACGACCTCATCGGTGGCAAACTGTCTAGCCTGAATCAG 2788  
 Qy 2820 GATTTGACGCTTCAGCCCTATCAGGTTCATGTGGCTCGAAATGCGCTGACGACGCTTCCC 2879  
 Db 2789 GATTTGACGCTTCAGCCCTATCAGGTTCATGTGGCTCGAAATGCGCTGACGACGCTTCCC 2848  
 Qy 2880 AATATCCGCTCAACGCTTTCAGACGGCATTTGGG 2914  
 Db 2849 AATATCCGCTCAACGCTTTCAGACGGCATTTGGG 2883

RESULT 4  
 ID AAT11179 standard; DNA; 2883 BP.  
 AC AAT11179;  
 DT 20-JUN-1996 (first entry)  
 XX Neisseria polysaccharea amylsucrase gene.  
 DE Amylosucrase; bacteria; fungi; plants; detection; transformation;  
 KW linear; alpha-1,4-glucans; amylose; sucrose; colourless;  
 KW odourless; tasteless; non-toxic; biodegradable; self-sustaining;  
 KW films; fibres; textiles; paper-making; glass-fibre; tablet binder;  
 KW food thickener; sound proofing; flow properties; paraffin oils;  
 KW organic compound inclusion; chromatographic separation;  
 KW cyclodextrins; ss.  
 XX OS Neisseria polysaccharea.  
 XX FH Key Location/Qualifiers  
 PT CDS 939..2783  
 FT /\*tag= a  
 XX DB4417879-AI.  
 PN 23-NOV-1995.  
 XX 18-MAY-1994; 94DE-4417879.  
 XX 18-MAY-1994; 94DE-4417879.  
 XX (GENB-) INST GENBIOLOGISCHE FORSCHUNG.  
 PI Buettcher V, Kossmann J, Welsh T;  
 XX WPI: 1996-000447/01.  
 DR P-PSDB; AAR8633.  
 XX New DNA sequence encoding amyl:sucrase of Neisseria - and  
 PT transformed plant, bacteria and fungi able to produce linear  
 PT alpha-1,4-glucan(s), esp. amylose, in practically pure form  
 XX Claim 6; Pages 27-33; 42pp; German.  
 XX The N. polysaccharea DNA sequence AAT11179, which encodes AAR8633  
 CC amylsucrase (ASA), can be used to produce bacteria, fungi and  
 CC plants that express ASA, and to detect and isolate related DNA  
 CC from other organisms. Transformed plants which express ASA are  
 CC able to produce linear alpha-1,4-glucans, specifically amylose  
 CC from sucrose, which can be used to produce colourless,  
 CC odourless, tasteless, nontoxic, biodegradable, self-sustaining  
 CC films or fibres, e.g. for use in the food, textile, paper-  
 CC making and glass-fibre industries. Amylose can also be used as  
 CC a binder for tablets, thickener for food, in sound proofing,  
 CC panels, to improve flow properties in paraffin-based oils, for  
 CC inclusion of organic cpds. in chromatographic sepn. and as a  
 CC starting material for cyclodextrins.  
 XX Sequence 2883 BP; 661 A; 852 C; 758 G; 612 T; 0 other;  
 SQ Query Match 90.6%; Score 2641; DB 17; Length 2883;

Best Local Similarity 97.8%; Pred. No. 0;  
 Matches 2852; Conservative 0; Mismatches 30; Indels 33; Gaps 16;  
 Qy 1 QAGTTTGGCGTTCCCGAACCAAGCTGATGCTTTCAGCGCGAACACCTGTCCGCGAAGCGG 60  
 Db 1 GAGTTTGGCGTTCCCGAACCAAGCTGATGCTTTCAGCGCGAACACCTGTCCGCGAAGCGG 60  
 Qy 61 CTGACCGCCCTTTTGGCCCATCGACATGTAATCGGTTTGGTGGGAGCTCTTTC 120  
 Db 61 -----ACCGCCCTTTTGGCCCATCGACATGTAATCGGTTTGGTGGGAGCTCTTTC 114  
 Qy 121 GCTTTGAGCGTGGCAGAAAGCAAGTTCAGACGCTTTCGCGCTTTTGGCGCATCACGCA 180  
 Db 115 GCTTTGAGCGTGGCAGAAAGCAAGTTCAGACGCTTTCGCG-GCCTTTGGCGCATCACGCA 173  
 Qy 181 ATTTTGAGATGTCCGCGCCGACGCTCTCCATCTGTTTTCAGACGGCATACGATTTCTCT 240  
 Db 174 ATTTTGAGATGTCCGCGCCGACGCTCTCCATCTGTTTTCAGACGGCATACGATTTCTCT 233  
 Qy 241 TCGCGCGCGTGGTGAACCTCATGTTTCAGACGAGGGCGGCGATGCGCTTTTGA 300  
 Db 234 TCGCGCGCGTGGTGAACCTCATGTTTCAGACGAG--GGCGATGCCGTTTGA 290  
 Qy 301 GCATGCGCGCAGCGCGCGCGGCTTTCGCGGAAAGCTCGATATCGATAATG 360  
 Db 291 GCATGCGCGCAGCGCGCGG-----AGCGTTTCGCGGAAAGCTCGATAATG 344  
 Qy 361 TCGGCGAGCGCGCTTTCATCAGGAGTTCGAGCTTCAAAATATATATATATATATATATAT 420  
 Db 345 TCGGCGAGCGCGCTTTCATCAGGAGTTCGAGCTTCAAAATATATATATATATATATATAT 404  
 Qy 421 GGGAAACAGCGCGCTTCGCCATGCCGCTCTGAACCTAAACAGCAGCGGCTTTCGCGGAGC 480  
 Db 405 GGGAAACAGCGCGCTTCGCCATGCCGCTCTGAACCTAAACAGCAGCGGCTTTCGCGGAGC 464  
 Qy 481 CGCTCGCGAGCGTCTCGGTGGCGCAATATCTTCGCCGATGCTCGCGCGCATTCGAAA 540  
 Db 465 CGCTCGCGAGCGTCTCGGTGGCGCAATATCTTCGCCGATGCTCGCGCGCATTCGAAA 524  
 Qy 541 AATTCGCGCGGAACCTCGACGATTCGAAAGCGGAGCTTTTGAATTTGGTCAAGTACGCG 600  
 Db 525 AATTCGCGCGGAACCTCGACGATTCGAAAGCGGAGCTTTTGAATTTGGTCAAGTACGCG 584  
 Qy 601 GAAAGTACGCGCATCGCGGCGACAAAGCGGCGAGTTCGCGGCTTCGCGCTTCG 660  
 Db 585 GAAAGTACGCGCATCGCGGCGACAAAGCGGCGAGTTCGCGGCTTCGCGCTTCG 644  
 Qy 661 ATACGCTGTTTTCAGGCTCA-GGCTGGTGTGCATGCGCGCTTCGCGCTGAAGAA 719  
 Db 645 ATACGCTGTTTTCAGGCTCAAGGCTGGTGTGCATGCGCGCTTCGCGCTGAAGAA 704  
 Qy 720 CGGTAAAGACGCAATTATAGCAAGGCGACAGGCAATGTTTCAGACGCGCATTTCTGTGCG 779  
 Db 705 CGGTAAAGACGCAATTATAGCAAGGCGACAGGCAATGTTTCAGACGCGCATTTCTGTGCG 764  
 Qy 780 CCGGCTGTATATGATATACAGCAGCATCCGATATCGGAATGCGACTTGGCAGACGCCCT 839  
 Db 765 CCGGCTGTATATGATATACAGCAGCATCCGATATCGGAATGCGACTTGGCAGACGCCCT 823  
 Qy 840 GTCTTTTCTAGTCACTCGGAGTTCCTTCAGTATGATTGCGACGACAGCGCTTACAGGCA 899  
 Db 824 GTCTTTTCTAGTCACTCGGAGTTCCTTCAGTATGATTGCGACGACAGCGCTTACAGGCA 883  
 Qy 900 TTTTCAGATACGCGCGCAGACCGCCGGTGGAACTTCAGAAATCGGAGCGAGGATCATG 959  
 Db 884 TTTTCAGATACGCGCGCAGACCGC--GTGCGAAACTTCAGAAATCGGAGCGAGGATCATG 941  
 Qy 960 TTGACCCCGCAGCGAGGCTCGGTTTTCATTTTACAGTACCTTCAAAAACAGCATCTTGAC 1019  
 Db 942 TTGACCCCGCAGCGAGGCTCGGTTTTCATTTTACAGTACCTTCAAAAACAGCATCTTGAC 1001  
 Qy 1020 ATCTACGCGCGCGAAGCGCGCGCATCGAAAATCGGAAGACTGGCGGAGTTTCG 1079

Db	1002	ATCTACAGCCCGAA	CHGCGCGCGCGGCAATCGAAAAATCGAAGACTGGCGGCGAGTTTTTCG	1061
Qy	1080	CGCGCATGATACGA	TTTTCCCAAACTGATGAACGAACCTCGACAGCGGTGTACGGCAAC	1139
Db	1062	CGCGCATGATACGA	TTTTCCCAAACTGATGAACGAACCTCGACAGCGGTGTACGGCAAC	1121
Qy	1140	AACGAAGCCCTGCT	CGCTATGCTGGAAATGCTGCTGGCGCAGGCGATGGCAAGCTATTCC	1199
Db	1122	AACGAAGCCCTGCT	CGCTATGCTGGAAATGCTGCTGGCGCAGGCGATGGCAAGCTATTCC	1181
Qy	1200	CAACGCAACTCATC	TTTAAAGATATCGATATCGCGCGGAAACAAACCCCGATTGGATT	1259
Db	1182	CAACGCAACTCATC	TTTAAAGATATCGATATCGCGCGGAAACAAACCCCGATTGGATT	1241
Qy	1260	TTGTCCAAACAA	CAAGTCGGCGCGGTGTGTAAGTTGTTGTGGCGCGAATTTGAAG	1319
Db	1242	TTGTCCAAACAA	CAAGTCGGCGCGGTGTGTAAGTTGTTGTGGCGCGAATTTGAAG	1301
Qy	1320	GGCTTGAAAGATAA	ATTCCTTATTTCAAGAGCTGGTTTGACTTATCTGCACCTGATG	1379
Db	1302	GGCTTGAAAGATAA	ATTCCTTATTTCAAGAGCTGGTTTGACTTATCTGCACCTGATG	1361
Qy	1380	CCGCTGTTTAAAT	GCCTGGAAGCAAAAGCGACGGCGGCTATCGGTTCAGCACTACCGC	1439
Db	1362	CCGCTGTTTAAAT	GCCTGGAAGCAAAAGCGACGGCGGCTATCGGTTCAGCACTACCGC	1421
Qy	1440	GATGTCAATCGG	CATGCGCACAATAGGCGACTTGGCGAAGCTCATTTGCTGGCTGCAC	1499
Db	1422	GATGTCAATCGG	CATGCGCACAATAGGCGACTTGGCGAAGCTCATTTGCTGGCTGCAC	1481
Qy	1500	GAACCGCGGCA	TTTTCCGCGCTGTCGATTTTATCTTCAACCAACCTCCACAGCAACAGAA	1559
Db	1482	GAATC--GCAT	TTTCCGCGCTGTCGATTTTATCTTCAACCAACCTCCACAGCAACAGAA	1539
Qy	1560	TGGCGCAACGCT	TCGGCGCGCGCGCGGACCCGCTTTTGGACHATTTCTACTATATTTTCCC	1619
Db	1540	T--GGCGCAACG	CTG--CGCGCGGACCCGCTTTTGGACHATTTCTACTATATTTTCCC	1595
Qy	1620	GACCGCGGATG	CCGCAACATACAGCCGACCCCTGCGCGAAATCTTCCCGGACCAAGCAC	1679
Db	1596	GACCGCGGATG	CCGCAACATACAGCCGACCCCTGCGCGAAATCTTCCCGGACCAAGCAC	1655
Qy	1680	CGCGCGGCTT	TCGCAACTGGAAGACGAGCGTGGGTGTGGACGACTTCAATTTCTTC	1739
Db	1656	CCGCGCGGCTT	TCGCAACTGGAAGACGAGCGTGGGTGTGGACGACTTCAATTTCTTC	1715
Qy	1740	CAATGGACATTGA	ATTACAGCACCCCTGGGTATTCGCGCAATGGCGGCGGAAATGCTG	1799
Db	1716	CAATGGACATTGA	ATTACAGCACCCCTGGGTATT--CGCGAATGGCGGCGGAAATGCTG	1774
Qy	1800	TTCCCTGCCAA	CTTGGCGGTTCATCTCGTATGGATCGGTTCGCTTTATTTGGAAA	1859
Db	1775	TTCCCTGCCAA	CTTGGCGGTTCATCTCGTATGGATCGGTTCGCTTTATTTGGAAA	1834
Qy	1860	CAATGGGACAG	CTCGGAAACCTTCGCGCAGGCGCACGCCCTTCATCGCGCGGTTCAT	1919
Db	1835	CAATGGGACAG	CTCGGAAACCTTCG--GCAGGCAAGCCCTTCATCGCGCGGTTCAT	1892
Qy	1920	GCCGTTATGGT	ATTGCGCGCCCGCGTGTCTTCAAATCGAAGCCATCGTCCACCCC	1979
Db	1893	GCCGTTATGGT	ATTGCGCGCCCGCGTGTCTTCAAATCGAAGCCATCGTCCACCCC	1952
Qy	1980	GACCAAGTGTG	CTCAATCATCGGCGAGACGAATGCGCAATTCGGTTACAACCCCTGCAA	2039
Db	1953	GACCAAGTGTG	CTCAATCATCGGCGAGAGAAATGCGCAATTCGGTTACAACCCCTGCAA	2012
Qy	2040	ATGSCATTGTT	GTGGAAACACCCTTGCACGGCGAAGTCAACTGTCTTCATCAGGCGCTG	2099
Db	2013	ATGSCATTGTT	GTGGAAACACCCTTGCACGGCGAAGTCAACTGTCTTCATCAGGCGCTG	2072
Qy	2100	ACCTACCGCA	AACTGCGCGACATACCGCTTGGGTTCAACTAGTTCGCGACGACAGAC	2159
Db	2073	ACCTACCGCA	AACTGCGCGACATACCGCTTGGGTTCAACTAGTTCGCGACGACAGAC	2132

RESULT 5	
AAF61709	
ID	AAF61709 standard; DNA; 1939 BP.
XX	
XX	
AC	AAF61709;
XX	
DT	12-JUL-2001 (first entry)
XX	
DE	Amylosucrase PCR derived DNA fragment SEQ ID 4.
XX	
XX	
KW	Amylosucrase; EC 2.4.1.4; fusion protein; GST; glutathione-S-transferase;
KW	poly(1,4-alpha-glucan); film production; food additive; cyclodextrin; ds.
XX	
XX	
QS	Unidentified.
XX	
FN	WO200125449-A2.
XX	
PD	12-APR-2001.
XX	

PF 04-OCT-2000; 2000WO-EP09695.  
XX 07-OCT-1999; 99DE-1048408.  
XX (AXIV-) AXIVA GMBH.  
XX Bengs H, Polakowski T, Held A, Gallert K;  
XX WPI; 2001-328330/34.  
DR Amylosucrase immobilized as fusion protein with anchoring sequence,  
XX useful in continuous preparation of poly(1,4-alpha-glucan) from sucrose  
XX .  
XX Claim 3; Page 27-28; 38pp; German.  
XX This invention describes a novel amylosucrase (AS), immobilizable on a  
XX solid phase, which comprises a fusion protein (FP) of functional units  
XX of AS, an anchoring sequence, and optionally additional auxiliary  
XX sequences. The invention also describes (1) nucleic acid (I) encoding  
XX FP; (2) expression vector containing (I) and able to express FP in a  
XX host cell; Escherichia coli containing the vector of (2); (3) anchoring  
XX sequence, or its functional variants or fragments, of at least 8  
XX nucleotides that encodes an epitope, a high-affinity binding partner or  
XX GST (Glutathione-S-transferase); (4) solid phase for immobilizing AS  
XX comprising Glutathione-Sepharose; (5) combined, stable catalyst (A)  
XX comprising FP immobilized on Sepharose for production of  
XX poly(1,4-alpha-glucan) (II); and (6) biocatalytic production of (II).  
XX AS is used in production of poly(1,4-alpha-glucans) (II), useful for  
XX producing films, as food additives, as starting materials for  
XX cyclodextrins and as auxiliaries in pharmaceutical formulations.  
XX Immobilized AS makes possible efficient, inexpensive and continuous  
XX production of poly(1,4-alpha-glucans) (II), and it can be used  
XX repeatedly. Compared with known methods, specificity is improved  
XX (increased yield of (II) and reduced formation of palatinose) and  
XX reaction is complete within 24 hours, compared to 48-72 hours for batch  
XX methods. This sequence encodes a PCR derived amylosucrase (EC 2.4.1.4)  
XX fragment described in the invention.  
XX Sequence 1939 BP; 453 A; 605 C; 482 G; 395 T; 4 other;  
XX  
XX Query Match 55.9%; Score 1921.4; DB 22; Length 1939;  
XX Best Local Similarity 99.9%; Field No. 0;  
XX Matches 1922; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
XX  
XX 955 TCATGTGACCCCAACGAGCAAGTCGGTTTGATTTTACAGTACTCAAAACAGCATCT 1014  
XX 10 TCATGTGACCCCAACGAGCAAGTCGGTTTGATTTTACAGTACTCAAAACAGCATCT 69  
XX  
XX 1015 TGGACATCTACAGCGCCGAAACGCGCGCGGCGATCGAAAATCCGAAGACTGGCGGAGT 1074  
XX 70 TGGACATCTACAGCGCCGAAACGCGCGCGGCGATCGAAAATCCGAAGACTGGCGGAGT 129  
XX  
XX 1075 TTTCGCGCGCATGATACGCAATTTCCCAACTGATGAACGAACTCGACAGCGGTGTACG 1134  
XX 130 TTTCGCGCGCATGATACGCAATTTCCCAACTGATGAACGAACTCGACAGCGGTGTACG 189  
XX  
XX 1135 GCAACAAAGAACGCTGTGCTGTGCTGTGGAATGCTGCTGGCGGAGCGATGGCAAGCT 1194  
XX 190 GCAACAAAGAACGCTGTGCTGTGGAATGCTGCTGGCGGAGCGATGGCAAGCT 249  
XX  
XX 1195 ATTCCCAACGCACTCTCCTTAAGATATCGATATCGCGCGGAAACCAACCCCGATT 1254  
XX 250 ATTCCCAACGCACTCTCCTTAAGATATCGATATCGCGCGGAAACCAACCCCGATT 309  
XX  
XX 1255 GGATTTTGTCCCAACAAAGTCGCGCGGTGTGCTGTGATTTGTTTTCGCGCGATT 1314  
XX 310 GGATTTTGTCCCAACAAAGTCGCGCGGTGTGCTGTGATTTGTTTTCGCGCGATT 369  
XX  
XX 1315 TCAAGGCTTGAAGATATAATTCCTTATTTCAAGAGCTTGGTTTGACTTATCTGCACC 1374  
XX 370 TCAAGGCTTGAAGATATAATTCCTTATTTCAAGAGCTTGGTTTGACTTATCTGCACC 429  
XX  
QY 1375 TGATGCGCGCTGTTTAATGCGCTGAGGCAAAAGCGAGCGGCGCTATGCGTTCAGCAGCT 1434  
DB 430 TGATGCGCGCTGTTTAATGCGCTGAGGCAAAAGCGAGCGGCGCTATGCGTTCAGCAGCT 489  
QY 1435 ACCGCGATGTCAATTCGCGCACTGCGGCAATAGCGCACTTCGCGGAAGTCATTCCTCGC 1494  
DB 490 ACCGCGATGTCAATTCGCGCACTGCGGCAATAGCGCACTTCGCGGAAGTCATTCCTCGC 549  
QY 1495 TGCACGAAGCGCGCATTTTCGCGCTGCTGCTGATTTTATCTTCAACCAACCTCCACCAAC 1554  
DB 550 TGCACGAAGCGCGCATTTTCGCGCTGCTGCTGATTTTATCTTCAACCAACCTCCACCAAC 609  
QY 1555 ACGAATGCGCGCAACGCTGCGCGCGCGCGCACTTCGCAATTTTCTACTATATTT 1614  
DB 610 ACGAATGCGCGCAACGCTGCGCGCGCGCGCACTTCGCAATTTTCTACTATATTT 669  
QY 1615 TCCCGGACCGCGGATGCGCGCAATAGCGCACTTCGCGCACTTCGCGGAATCTTCCCGGACC 1674  
DB 670 TCCCGGACCGCGGATGCGCGCAATAGCGCACTTCGCGCACTTCGCGGAATCTTCCCGGACC 729  
QY 1675 AGCACCGCGCGGCTTCTCGCACTGGAAGCAAGCACTGCGGTGGAACCACTTCAAT 1734  
DB 730 AGCACCGCGCGGCTTCTCGCACTGGAAGCAAGCACTGCGGTGGAACCACTTCAAT 789  
QY 1735 CTTTCCAAATGGGACTTGAATTAAGCAACCGGTGGGTATTCGCGCAATAGCGCGGCAAA 1794  
DB 790 CTTTCCAAATGGGACTTGAATTAAGCAACCGGTGGGTATTCGCGCAATAGCGCGGCAAA 849  
QY 1795 TGTGTTTCTTGCACCTGCGGCTTGACATCTGCGTATGATGCGGTGTCCTTTATTT 1854  
DB 850 TGTGTTTCTTGCACCTGCGGCTTGACATCTGCGTATGATGCGGTGTCCTTTATTT 909  
QY 1855 GGAACAAATGGGCAAGCTGCGGAAACCTGCGGAGGCGACGCGCTCATTCGCGCGGT 1914  
DB 910 GGAACAAATGGGCAAGCTGCGGAAACCTGCGGAGGCGACGCGCTCATTCGCGCGGT 969  
QY 1915 TCATGCGGTATGCGTATTCGCGCGCGCGCGGTGTTCTTCAATCCGAGGCACTTCGTC 1974  
DB 970 TCATGCGGTATGCGTATTCGCGCGCGCGCGGTGTTCTTCAATCCGAGGCACTTCGTC 1029  
QY 1975 ACCCGGACCAAGTCGTCCTCAATATACATCGGCGAGGACGAATGCAATCGGTTTACACCC 2034  
DB 1030 ACCCGGACCAAGTCGTCCTCAATATACATCGGCGAGGACGAATGCAATCGGTTTACACCC 1089  
QY 2035 TGAATATGGCACTGTTTGTGAACACCTTTCGACGCGCGGAGTCAACTGCTCCATCAGG 2094  
DB 1090 TGAATATGGCACTGTTTGTGAACACCTTTCGACGCGCGGAGTCAACTGCTCCATCAGG 1149  
QY 2095 CGCTGACCTACCGCCCAACCTGCGCGGAGCATACCGCTGCGGTCAAATACTGTCGCGAGCC 2154  
DB 1150 CGCTGACCTACCGCCCAACCTGCGCGGAGCATACCGCTGCGGTCAAATACTGTCGCGAGCC 1209  
QY 2155 ACGACGACATCGGCTGGAGCTTTCGCGATGAAGCGGCGATATCTGGGCAATAGCGGCT 2214  
DB 1210 ACGACGACATCGGCTGGAGCTTTCGCGATGAAGCGGCGATATCTGGGCAATAGCGGCT 1269  
QY 2215 ACGACGACATCGGCTGGAGCTTTCGCGATGAAGCGGCGATATCTGGGCAATAGCGGCT 2274  
DB 1270 ACGACGACATCGGCTGGAGCTTTCGCGATGAAGCGGCGATATCTGGGCAATAGCGGCT 1329  
QY 2275 GTGCGGTACCGTTTCCAAATACAAACCAAGCAAGCGAGCTGCGGTGTCAGTGGTACAGCG 2334  
DB 1330 GTGCGGTACCGTTTCCAAATACAAACCAAGCAAGCGAGCTGCGGTGTCAGTGGTACAGCG 1389  
QY 2335 CGGCAATGTTGCGCTTGGCGGAGAGCATCCCGACGCGGTGACCGGATCAAACTCTTGT 2394  
DB 1390 CGGCAATGTTGCGCTTGGCGGAGAGCATCCCGACGCGGTGACCGGATCAAACTCTTGT 1449  
QY 2395 ACAGCATTTGCTTTGAGTACCGCGGCTCTGCGCTGATTTTACCTAGGCGGACGAGTGGGTA 2454  
DB 1450 ACAGCATTTGCTTTGAGTACCGCGGCTCTGCGCTGATTTTACCTAGGCGGACGAGTGGGTA 1509  
QY 2455 CGCTCAATGACGAGCATGTTGTCGCAAGACGAGCAATAGAGCGGACGAGCGGCTTGGGCGC 2514

Db	1510	CGCTCAATGACGACGACTGGTTCGCACACAGCAATTAAGCGACGACGACGCTGTGGGGCGC	1569
QY	2515	ACCGTTCGGGCTACAAAGAGCCCTGTACGCGCAAAGCAACGATCCGTTCACCGCAGCGCG	2574
Db	1570	ACCGTTCGGGCTACAAAGAGCCCTGTACGCGCAAAGCAACGATCCGTTCACCGCAGCGCG	1629
QY	2575	GGCAATCTATCAGGGCTTCGGGCATATGATGCGTCCGCCAAAGCAATCCGGGCTTCG	2634
Db	1630	GGCAATCTATCAGGGCTTCGGGCATATGATGCGTCCGCCAAAGCAATCCGGGCTTCG	1689
QY	2635	ACGGCGCAGCGCTGGTTACATTCAAACACCAACCAAGCAGCATCATCGGCTACATCCGCA	2694
Db	1690	ACGGCGCAGCGCTGGTTACATTCAAACACCAACCAAGCAGCATCATCGGCTACATCCGCA	1749
QY	2695	ACAATCGCTTTTGGCAATTCGGTAATTCAGCGCAATTCGCAAAACGGTTACCGCGCAT	2754
Db	1750	ACAATCGCTTTTGGCAATTCGGTAATTCAGCGCAATTCGCAAAACGGTTACCGCGCAT	1809
QY	2755	CCCTGCAAGCCATCCCTTCAAGCGCAGCAACCTCATCGGTGGCAAACTGTACAGCCTGA	2814
Db	1810	CCCTGCAAGCCATCCCTTCAAGCGCAGCAACCTCATCGGTGGCAAACTGTACAGCCTGA	1869
QY	2815	ATCAGATTTGACGCTTCAGCCCTATCAGGTTCATGTGGCTCGAAATCGCTGACGACGC	2874
Db	1870	ATCAGATTTGACGCTTCAGCCCTATCAGGTTCATGTGGCTCGAAATCGCTGACGACGC	1929
QY	2875	TTT 2877	
Db	1930	CTC 1932	

RESULT 6

AAAF61711	ID	AAAF61711 standard; DNA; 6878 BP.
XX	AC	AAAF61711;
XX	AC	
XX	AC	
DT		12-JUL-2001 (first entry)
XX		
DE		Expression vector pGEX-4T-1-RMSU containing amylosucrase DNA.
XX		
KW		Amylosucrase; EC 2.4.1.4; fusion protein; GST; glutathione-S-transferase;
KW		poly(1,4-alpha-glucan); film production; food additive; cyclodextrin; ds.
XX		
OS		Unidentified.
XX		
PN		WO200125449-A2.
XX		
PD		12-APR-2001.
XX		
PF		04-OCT-2000; 2000WO-EP09695.
XX		
XX		07-OCT-1999; 99DE-1048408.
PR		
XX		(AXIV-) AXIVA GMBH.
PA		
XX		
PI		Bengs H, Polakowski T, Held A, Gallert K;
XX		
DR		WPI; 2001-328330/34.
XX		
PT		Amylosucrase immobilized as fusion protein with anchoring sequence,
PT		useful in continuous preparation of poly(1,4-alpha-glucan) from sucrose
XX		
PS		Example 2: Page 29-31; 38pp; German.
XX		
CC		This invention describes a novel amylosucrase (AS), immobilizable on a
CC		solid phase, which comprises a fusion protein (FP) of functional units
CC		of AS, an anchoring sequence, and optionally additional auxiliary
CC		sequences. The invention also describes (1) nucleic acid (I) encoding
CC		poly; (2) expression vector containing (I) and able to express FP in a
CC		host cell; Escherichia coli containing the vector of (2); (3) anchoring

1663 AGCACCGGGGGGTTCTTCGCAACTGGAGAGCGGACGCTGGGTGFGAGACGACCTTCAATT 1722  
1735 CTTTCCAAATGGGACTTGAATTACAGCAACCGGTGGGTATTCGGGCGCAATGGCGGCGGAAA 1794  
1723 CTTTCCAAATGGGACTTGAATTACAGCAACCGGTGGGTATTCGGGCGCAATGGCGGCGGAAA 1782  
1795 TGTGTTCTTTCGCAACTGGGCGGTGTCATCTCTGGGTATGATGCGGTTCCTTTATTT 1854  
1783 TGTGTTCTTTCGCAACTGGGCGGTGTCATCTCTGGGTATGATGCGGTTCCTTTATTT 1842  
1855 GGAACCAATGGGAGCGGAGCTGCGAARACCTGCGGAGCGGACGCGCTCATTCGCGGGT 1914  
1843 GGAACCAATGGGAGCGGAGCTGCGAARACCTGCGGAGCGGACGCGCTCATTCGCGGGT 1902  
1915 TCAATGCGGTTATGCGTATTCGCGGCGCGCGGTGTTCTTCAAAATCGAAGCCATCGTCC 1974  
1903 TCAATGCGGTTATGCGTATTCGCGGCGCGCGGTGTTCTTCAAAATCGAAGCCATCGTCC 1962  
1975 ACCCGGACCAAGTGTCAATACATCGGGGAGGAGGATGCAATGGTTACACCCCC 2034  
1963 ACCCGGACCAAGTGTCAATACATCGGGGAGGAGGATGCAATGGTTACACCCCC 2022  
2035 TGCAAATGGCAATGTTGTGGGAACACCTTGCACGCGGGAAGTCAACCTGCTCCATCAGG 2094  
2023 TGCAAATGGCAATGTTGTGGGAACACCTTGCACGCGGGAAGTCAACCTGCTCCATCAGG 2082  
2095 CGTGACCTACGCGCACAACTGCGGAGGATACCGCGTGGGTCACTACGTCGCGAGCC 2154  
2083 CGTGACCTACGCGCACAACTGCGGAGGATACCGCGTGGGTCACTACGTCGCGAGCC 2142  
2155 ACGAGCAGATCGGCTGGGAGCTTTCGCGATGAGAGCGGCGCATATCTGGGCGATAGCGGCT 2214  
2143 ACGAGCAGATCGGCTGGGAGCTTTCGCGATGAGAGCGGCGCATATCTGGGCGATAGCGGCT 2202  
2215 ACGAGCAGCAGCAATTCCTCAACCGCTTCTTGTCACACCGTTTCGACGGGAGCTTCGCTC 2274  
2203 ACGAGCAGCAGCAATTCCTCAACCGCTTCTTGTCACACCGTTTCGACGGGAGCTTCGCTC 2262  
2275 GTGGGATACGCTTCAATACACCAACGACGACGAGGAGCTGCGGTGTCAGTGGTACAGCGG 2334  
2263 GTGGGATACGCTTCAATACACCAACGACGACGAGGAGCTGCGGTGTCAGTGGTACAGCGG 2322  
2335 CGGCATTTGCTGGCTTGGCGAAGACGATCCCGACGCGTTTGACCGCATCAAACTCTTGT 2394  
2323 CGGCATTTGCTGGCTTGGCGAAGACGATCCCGACGCGTTTGACCGCATCAAACTCTTGT 2382  
2395 ACAGCATTTGCTTGGTACGCGGCTGCTGCGCTGATTTACTAGCGGACGAGTGGGTA 2454  
2383 ACAGCATTTGCTTGGTACGCGGCTGCTGCGCTGATTTACTAGCGGACGAGTGGGTA 2442  
2455 CGCTCAATGACGACGACTGGTTCGCAAGACAGCAATAGAGGACGACGACGCGTTGGGCGC 2514  
2443 CGCTCAATGACGACGACTGGTTCGCAAGACAGCAATAGAGGACGACGACGCGTTGGGCGC 2502  
2515 ACCGTCCGCGCTTACACGAGCCCTGTAGCGCAACGCAACGATCCGTGACCGGAGCGG 2574  
2503 ACCGTCCGCGCTTACACGAGCCCTGTAGCGCAACGCAACGATCCGTGACCGGAGCGG 2562  
2575 GGCATATTCAGGCTTGGCGCATATGATTTGCGCTCGCCAAAGCAATCGGCTTCG 2634  
2563 GGCATATTCAGGCTTGGCGCATATGATTTGCGCTCGCCAAAGCAATCGGCTTCG 2622  
2635 ACGGCGGAGGCTGTTTACATTCACACCAACCAAGCACATCTCGCTACATCCGCA 2694  
2623 ACGGCGGAGGCTGTTTACATTCACACCAACCAAGCACATCTCGCTACATCCGCA 2682  
2695 ACAATGCGCTTTGGATTCGCTTACCTTACGAGGATATTCGCAACCGTTACCGGCGATA 2754  
2683 ACAATGCGCTTTGGATTCGCTTACCTTACGAGGATATTCGCAACCGTTACCGGCGATA 2742  
2755 CCTCGAAGCCATGCCCTTCAAGGGGCGACGACTCATCGTGGGCAAACTGTCAGCTGTA 2814

2743 CCTGCAAGCCATGCGCTTCAAGGGGACGACCTCATCTGCTGGCGCAAACTGTCAGCCTGA 2802  
2815 ATCAGATTTGAGCGCTTACGCGCTATCAGGTATGAGCTATGAGCTCGAAATCGCTGACGACGC 2874  
2803 ATCAGATTTGAGCGCTTACGCGCTATCAGGTATGAGCTATGAGCTCGAAATCGCTGACGACGC 2862  
2875 TTC 2877  
2863 CTC 2865  
RESULT 7  
AAF61710  
ID AAF61710 standard; DNA; 1910 BP.  
XX  
AC AAF61710;  
XX  
DT 12-JUL-2001 (first entry)  
XX  
DB Amylosucrase PCR derived DNA fragment AMSU5.  
XX  
KW Amylosucrase; EC 2.4.1.4; fusion protein; GST; glutathione-S-transferase;  
XX poly(1,4-alpha-glucan); film production; food additive; cyclodextrin; ds.  
XX  
OS Unidentified.  
XX  
PW WO200125449-A2.  
XX  
PD 12-APR-2001.  
XX  
PF 04-OCT-2000; 2000WO-EP09695.  
XX  
PR 07-OCT-1999; 99DB-1048408.  
XX  
PA (AXIV-) AXIVA GMEH.  
XX  
PI Bengs H, Polakowski T, Held A, Gallert K;  
XX  
DR MPI; 2001-328330/34.  
XX  
PT Amylosucrase immobilized as fusion protein with anchoring sequence.  
XX useful in continuous preparation of poly(1,4-alpha-glucan) from sucrose  
XX  
PS Claim 3; Page 28; 38pp; German.  
XX  
CC This invention describes a novel amylosucrase (AS), immobilizable on a  
CC solid phase, which comprises a fusion protein (FP) of functional units  
CC of AS, an anchoring sequence, and optionally additional auxiliary  
CC sequences. The invention also describes (i) nucleic acid (I) encoding  
CC FP; (2) expression vector containing (I) and able to express FP in a  
CC host cell; (3) Escherichia coli containing the vector of (2); (3) anchoring  
CC sequence, or its functional variants or fragments, of at least 8  
CC nucleotides that encodes an epitope, a high-affinity binding partner or  
CC GST (glutathione-S-transferase); (4) solid phase for immobilizing AS  
CC comprising glutathione-Sepharose; (5) combined, stable catalyst (A)  
CC comprising FP immobilized on Sepharose for production of (II).  
CC poly(1,4-alpha-glucan) (II); and (6) biocatalytic production of (II).  
CC AS is used in production of poly(1,4-alpha-glucans) (II), useful for  
CC producing films, as food additives, as starting materials for  
CC cyclodextrins and as auxiliaries in pharmaceutical formulations.  
CC Immobilized AS makes possible efficient, inexpensive and continuous  
CC production of poly(1,4-alpha-glucans) (II), and it can be used  
CC repeatedly. Compared with known methods, specificity is improved  
CC (increased yield of (II) and reduced formation of palatinose) and  
CC reaction is complete within 24 hours, compared to 48-72 hours for batch  
CC methods. This sequence encodes a PCR derived amylosucrase (BC 2.4.1.4)  
CC AMSU5 fragment described in the invention.  
XX  
SQ Sequence 1910 BP; 447 A; 595 C; 474 G; 387 T; 7 other;  
Query Match 64.9%; Score 1891.4; DB 22; Length 1910;  
Best Local Similarity 99.9%; Pred. No. 0;





FF	Key	Location/Qualifiers	DB	1186	TCGATATCGCGCGGAAAAAACAACCCGATGGATTTTGTCCAAACAACAGTCGCGCGG	1245
FF	CDS	258..2828	QY	1285	TGCTCACTGATGATTTGTTTTCGCGCGGATTTGAGAGGCTTGAAAGATAAAATTCCTTATT	1344
FF		/tag= a	DB	1246	TGCTCACTGATGATTTGTTTTCGCGCGGATTTGAGAGGCTTGAAAGATAAAATTCCTTATT	1305
FF		/product= "Amsus"	QY	1345	TTCAAGAGCTTTGGTTTGACCTTATCTGCACCTGATGCGCGCTTTTAAATTCCTTGAAGCA	1404
FF	WO200125449-A2.		DB	1306	TTCAAGAGCTTTGGTTTGACCTTATCTGCACCTGATGCGCGCTTTTAAATTCCTTGAAGCA	1365
FF	12-APR-2001.		QY	1405	AAAGCGACGGGCTATGCGGTGAGCGAGCTACCGGATGTCAATCCGCGACTGGGCAAA	1464
FF	04-OCT-2000; 2000WO-EP09695.		DB	1366	AAAGCGACGGGCTATGCGGTGAGCGAGCTACCGGATGTCAATCCGCGACTGGGCAAA	1425
FF	07-OCT-1999; 99DE-1048408.		QY	1465	TAGCGACTTTCGCGGAAAGTCAATTCCTGCGCTGCACGAAAGCGGCAATTTCCGCGCTGCG	1524
FF	(AXIV-) AXIVA GMBH.		DB	1426	TAGCGACTTTCGCGGAAAGTCAATTCCTGCGCTGCACGAAAGCGGCAATTTCCGCGCTGCG	1485
FF	Bengs H, Polakowski T, Held A, Gallert K;		QY	1525	ATTATATCTCAACACACCTTCCAAACGAAATGGGCGGACGCTGCGCGCGCGG	1584
FF	WPI; 2001-328330/34.		DB	1486	ATTATATCTCAACACACCTTCCAAACGAAATGGGCGGACGCTGCGCGCGCGG	1545
FF	P-PSDB; AA570883.		QY	1585	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1644
FF	Amylosucrase immobilized as fusion protein with anchoring sequence,		DB	1546	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	useful in continuous preparation of poly(1,4-alpha-glucan) from sucrose		QY	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	Claim 6; Page 31-34; 38pp; German.		QY	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	This invention describes a novel amylase (AS), immobilizable on a		DB	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	solid phase, which comprises a fusion protein (FP) of functional units		QY	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	of AS, an anchoring sequence, and optionally additional auxiliary		DB	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	sequences. The invention also describes (1) nucleic acid (I) encoding		QY	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	FP; (2) expression vector containing (I) and able to express FP in a		DB	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	host cell; Escherichia coli containing the vector of (2); (3) anchoring		QY	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	sequence, or its functional variants or fragments, of at least 8		DB	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	nucleotides that encodes an epitope, a high-affinity binding partner or		QY	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	GST (glutathione-S-transferase); (4) solid phase for immobilizing AS		DB	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	comprising glutathione-Sepharose; (5) combined, stable catalyst (A)		QY	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	comprising FP immobilized on Sepharose for production of (II).		DB	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	poly(1,4-alpha-glucan) (II); and (6) biocatalytic production of (II).		QY	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	AS is used in production of poly(1,4-alpha-glucans) (II), useful for		DB	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	producing films, as food additives, as starting materials for		QY	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	Cyclodextrins and as auxiliaries in pharmaceutical formulations.		DB	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	Immobilized AS makes possible efficient, inexpensive and continuous		QY	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	production of poly(1,4-alpha-glucans) (II), and it can be used		DB	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	repeatedly. Compared with known methods, specificity is improved		QY	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	(increased yield of (II) and reduced formation of palatinose) and		DB	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	reaction is complete within 24 hours, compared to 48-72 hours for batch		QY	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	methods. This sequence encodes the expression vector construct		DB	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	pGEX-4T-1-Amsus which contains the amylase (EC 2.4.1.4) described in		QY	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	the invention.		DB	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	Sequence 6851 BP; 1668 A; 1791 C; 1759 G; 1633 T; 0 other;		QY	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	Query Match 64.9%; Score 1891.4; DB 22; Length 6851;		DB	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	Best Local Similarity 99.9%; Pred. No. 0;		QY	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	Matches 1892; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		DB	1645	ACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCGACCAATACG	1605
FF	985 TGAATTTACAGTACCTCAAAACAGCATCTTGACATCTACAGCCGCAACAGCGCGCG	1044	QY	985	TGAATTTACAGTACCTCAAAACAGCATCTTGACATCTACAGCCGCAACAGCGCGCG	1044
FF	946 TGAATTTACAGTACCTCAAAACAGCATCTTGACATCTACAGCCGCAACAGCGCGCG	1005	DB	946	TGAATTTACAGTACCTCAAAACAGCATCTTGACATCTACAGCCGCAACAGCGCGCG	1005
FF	1045 GCATCGAAATTCGAGAGCTGCGCGAGTTTTCGCGCGGATGATAGCATTTCCCA	1104	QY	1045	GCATCGAAATTCGAGAGCTGCGCGAGTTTTCGCGCGGATGATAGCATTTCCCA	1104
FF	1006 GCATCGAAATTCGAGAGCTGCGCGAGTTTTCGCGCGGATGATAGCATTTCCCA	1065	DB	1006	GCATCGAAATTCGAGAGCTGCGCGAGTTTTCGCGCGGATGATAGCATTTCCCA	1065
FF	1105 AACTGATGACGACTCGACAGGCTGACGAGCAACAGCAAGCCCTGCTGCTATGCTGG	1164	QY	1105	AACTGATGACGACTCGACAGGCTGACGAGCAACAGCAAGCCCTGCTGCTATGCTGG	1164
FF	1066 AACTGATGACGACTCGACAGGCTGACGAGCAACAGCAAGCCCTGCTGCTATGCTGG	1125	DB	1066	AACTGATGACGACTCGACAGGCTGACGAGCAACAGCAAGCCCTGCTGCTATGCTGG	1125
FF	1165 AAATGCTGCTGGCGAGGATGCAAGATTTTCCCAAGCAACTCATCTTTAAAGATA	1224	QY	1165	AAATGCTGCTGGCGAGGATGCAAGATTTTCCCAAGCAACTCATCTTTAAAGATA	1224
FF	1126 AAATGCTGCTGGCGAGGATGCAAGATTTTCCCAAGCAACTCATCTTTAAAGATA	1185	DB	1126	AAATGCTGCTGGCGAGGATGCAAGATTTTCCCAAGCAACTCATCTTTAAAGATA	1185
FF	1225 TCGATATCGCGCGGAAAAACAACCCCGATTTGTTTTCACAAACAAGTCGCGCGG	1284	QY	1225	TCGATATCGCGCGGAAAAACAACCCCGATTTGTTTTCACAAACAAGTCGCGCGG	1284

Db 2266 CAGGCGACTGCGGTGTCAGTGGTACAGCCGCGGCATTTGGTGGCGCAAGCGATC 2325  
Qy 2365 CCACGCCGTTGACCCATCAAACTTTGTACAGCAATGCTTTGAGTACCGCGGTCTGC 2424  
Db 2326 CCACGCCGTTGACCCATCAAACTTTGTACAGCAATGCTTTGAGTACCGCGGTCTGC 2385  
Qy 2425 CGCTGATTTACCTTAGCGGAGTGGTACGCTCAATGACGAGCTGCTGCGCAAGACA 2484  
Db 2386 CGCTGATTTACCTTAGCGGAGTGGTACGCTCAATGACGAGCTGCTGCGCAAGACA 2445  
Qy 2485 GCAATAGAGCGACGACAGCCGTTGGCGCACCGTCCGCGCTACAAAGAGCCCTGTACG 2544  
Db 2446 GCAATAGAGCGACGACAGCCGTTGGCGCACCGTCCGCGCTACAAAGAGCCCTGTACG 2505  
Qy 2545 CGCAAGCGCAACGATCGGTGACGCGGCGCGGCAATCTATCAGGGCTTGCGCCATGCA 2604  
Db 2506 CGCAAGCGCAACGATCGGTGACGCGGCGCGGCAATCTATCAGGGCTTGCGCCATGCA 2565  
Qy 2605 TTCCCGTTCGCGCAAGCAATCCGCGCTTCGACGCGCGGCGAGGCTGGTTACATTCACACCA 2664  
Db 2566 TTCCCGTTCGCGCAAGCAATCCGCGCTTCGACGCGCGGCGAGGCTGGTTACATTCACACCA 2625  
Qy 2665 ACAACAGCAATCATCGGTGATCGGCGCAACATGCGCTTTTGGCATTCGGTAATCTCA 2724  
Db 2626 ACAACAGCAATCATCGGTGATCGGCGCAACATGCGCTTTTGGCATTCGGTAATCTCA 2685  
Qy 2725 GCGAATATCGCAAAACCGTTACCGCGCATACCTCGCAAGCCATCGCCCTTCAAGCGCACG 2784  
Db 2686 GCGAATATCGCAAAACCGTTACCGCGCATACCTCGCAAGCCATCGCCCTTCAAGCGCACG 2745  
Qy 2785 ACTCATGCTGGCGAAACTGTGAGCTGATCAGCTGATCAGGATTTGAGCGCTTCAGCCCTATCAGG 2844  
Db 2746 ACTCATGCTGGCGAAACTGTGAGCTGATCAGCTGATCAGGATTTGAGCGCTTCAGCCCTATCAGG 2805  
Qy 2845 TCATGCGCTCGAAATCGCTGAGCGACGCTTC 2877  
Db 2806 TCATGCGCTCGAAATCGCTGAGCGACGCTTC 2838

RESULT 9  
AAS97236/c  
ID AAS97236 standard; DNA; 795 BP.  
XX AC AAS97236;  
XX XX 12-MAR-2002 (first entry)  
XX DE Neisseria meningitidis virulence gene #41.  
XX KW Meningitis; virulence; gene; antibacterial; vaccine; veterinary;  
XX KW infection; Gram-negative bacteria; antimicrobial; ds.  
XX OS Neisseria meningitidis.  
XX XX WO200185772-A2.  
XX PD 15-NOV-2001.  
XX XX 08-MAY-2001; 2001WO-GB02003.  
XX PF 08-MAY-2000; 2000GB-0011108.  
XX PR 08-MAY-2000; 2000GB-0011108.  
XX XX (MICR-) MICROSCIENCE LTD.  
XX PA Tang C;  
XX FI WPI; 2002-066593/09.  
XX DR P-PSDB; AAU72951.  
XX XX

XX PT New peptide encoded by operon including virulence genes of Neisseria  
XX PT meningitidis, useful as vaccine component for treating or preventing  
XX PT meningitis and for identifying antimicrobial drug -  
XX XX

PS Claim 1; Page 174-175; 423pp; English.

XX The invention relates to a peptide (I) encoded by an operon (II) of  
CC Neisseria meningitidis including virulence genes, or a related molecule  
CC having a 40% sequence similarity at the peptide or nucleotide level in a  
CC Gram-negative bacterium, or its functional fragment, for therapeutic or  
CC diagnostic use. (I) and (II) are useful in the manufacture of a  
CC medicament for treating or preventing a condition (e.g., meningitis)  
CC associated with infection by Neisseria or Gram-negative bacteria. The  
CC product is useful for veterinary treatment and in a screening assay for  
CC the identification of an antimicrobial drug. The vaccines have  
CC prophylactic applications. AAS97196-AAS97305 represent N. meningitidis  
CC virulence genes and related PCR primers of the invention.

XX Sequence 795 BP; 166 A; 238 C; 223 G; 168 T; 0 other;

Query Match 22.5%; Score 655; DB 24; Length 795;  
Best Local Similarity 96.4%; Pred No. 5, 4e-170;  
Matches 670; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Qy 1 GAGTTTTCGGTTCCCGAACCGAAGCTGATGCTTGGAGCGAAGCACTGTCGCGCAAGCGG 60  
Db 728 GAGTTTTCGGTTCCCGAACCGAAGCTGATGCTTGGAGCGAAGCACTGTCGCGCAAGCGG 669  
Qy 61 CTGACGCGCCCTTTTGGCCCATCGACATCGTAACATCGGTTTGGTGGCAAGCTCTTTC 120  
Db 668 CTGACGCGCCCTTTTGGCCCATCGACATCGTAACATCGGTTTGGTGGCAAGCTCTTTC 609  
Qy 121 GCTTTGAGCGTGGCAGAAAGCAAGCTCAGACGCTTTCGCGCTTTGGCGCATCCCGCA 180  
Db 608 GCTTTGAGCGTGGCAGAAAGCAAGCTCAGACATCTTCGCGCTTTGGCGCATCCCGCA 549  
Qy 181 ATTTTGCAGATGTCGCGCGCGGCGAGTCTCCATCTGTTTCAGACGCGCATACGATTTCT 240  
Db 548 ATTTTGCAGATGTCGCGCGCGGCGAGTCTCCATCTGTTTCAGACGCGCATACGATTTCT 489  
Qy 241 TCGCGCGCGTTCGCTGAAACTCATGATTCAGACGAGCGCGCGGCGATGCCGTTTTTCA 300  
Db 488 TCGCGCGCGTTCGCTGAAACTCATGATTCAGACGAGCGCGCGGCGATGCCGTTTTTCA 429  
Qy 301 GCATCGCCACGCGCGCGCGGCGAGCGCGTTTCGCGCGGAAAGCTCGATTCGATG 360  
Db 428 GCATTTGCCACGCGCGCACCGGCGGTTTCGCGCGGAAAGCTCGATTCGATG 369  
Qy 361 TCGCGCGAGCGCGCTTTCAATCAGCGAGTTCGAGCGAGTTCAGAAATAATAATCGTCCGAACAC 420  
Db 368 TCGCGCGAGCGCGCTTTTCGATCAGCGCGTTCGAGCGATTCAGAAATAATAATCGTCCGAACAC 309  
Qy 421 GGGAAACGCGCGCTTTCGCGCATGCGCTCTGAACGTAACAGCGCGCTTTCGCGCGACG 480  
Db 308 GGGAAACGCGCGCTTTCGCGCATGCGCTCTGAACGTAACAGCGCGCTTTCGCGCGACG 249  
Qy 481 GCGTTCGCGAGCGCTTTCGCGCATGCGCTCTGAACGTAACAGCGCGCTTTCGCGCGACG 540  
Db 248 GCGTTCGCGAGCGCTTTCGCGCATGCGCTCTGAACGTAACAGCGCGCTTTCGCGCGACG 189  
Qy 541 AAATCGCGCGGAAACTTCGACGATTCGAGGCGAGGCTTTTCAATTTGGTCAAGTACGCGG 600  
Db 188 AAATCGCGCGGAAACTTCGACGATTCGAGGCGAGGCTTTTCAATTTGGTCAAGTACGCGG 129  
Qy 601 GAAAGTACGCGCGCATCGCGCGGCGAGCGCGCGGCGATTTTCGCGCGCTTCGCGTTCG 660  
Db 128 GAAAGTACGCGCGCATCGCGCGGCGAGCGCGCGGCGATTTTCGCGCGCTTCGCGTTCG 69  
Qy 661 ATAACGCGTGTTCGACGCTCAGGCTGGTGTGCAT 695  
Db 68 ATAACGCGTGTTCGACACACAGCGCGAGGACACAT 34

RESULT 10  
ABX09900/c  
ID ABX09900 standard; DNA; 765 BP.  
XX XX





CC Neisseria proteins given in AAB58550 to AAB58593, and AAB21589 to  
CC AAB21606 represent PCR primers which are used in the exemplification of  
CC the present invention. The NMB genome and fragments from it have  
CC antibacterial activity, and can be used in vaccines and gene therapy.  
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the  
CC proteins can be used in compositions for treating or preventing infection  
CC due to Neisseria bacteria or as a diagnostic reagent for detecting the  
CC presence of Neisseria bacteria or of antibodies raised to Neisseria  
CC bacteria. Computers, computer memory, computer storage medium or computer  
CC databases can be used in a search to identify open reading frames (ORFs)  
CC or coding sequences within the NMB genome. The DNA sequences provide  
CC further opportunities to find antigenic or immunogenic proteins which are  
CC more effective in vaccines than the outer membrane proteins currently  
CC used.

XX  
SQ Sequence 349980 BP; 86771 A; 92803 C; 86340 G; 84066 T; 0 other;  
Query Match 21.9%; Score 639; DB 21; Length 349980;  
Best Local Similarity 95.0%; Pred. No. 2.5e-164;  
Matches 660; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 1 GAGTTTGGCTCCCGACGACGATGATGCTTGAGCCGACACCTGTCCGCGAAGCGG 60  
DB GAGTTTGGCTCCCGACGACGATGATGCTTGAGCCGACACCTGTCCGCGAAGCGG 290445

QY 61 CTGACCGCCCTTTTGGCCCATCGACATCGTAACATCGGTTGGTGGCAAGCTCTTTC 120  
DB CTGACCGCCCTTTTGGCCCATCGACATCGTAACATCGGTTGGTGGCAAGCTCTTTC 290385

QY 121 GCTTTGAGCGTGGCGAAGCAAGTACGACGCGTCTTCCGCGCTTTGGCGATCACCGCA 180  
DB GCTTTGAGCGTGGCGAAGCAAGTACGACGCGTCTTCCGCGCTTTGGCGATCACCGCA 290325

QY 181 AFTTTGAGATGTCGCGCCGCGCTCTCTGTTTACAGCGCATACGATTTCTTCT 240  
DB AFTTTGAGATGTCGCGCCGCGCTCTCTGTTTACAGCGCATACGATTTCTTCT 290265

QY 241 TCGCGCGCGTGGTGAATCTATGATGTCAGACGAGCGGCGGATGTCGCTTTTGA 300  
DB TCGCGCGCGTGGTGAATCTATGATGTCAGACGAGCGGCGGATGTCGCTTTTGA 290205

QY 301 GCATGCGCCACGCGCGCGACGCGGTTTCGCGGGAAGCAAGCTCGATATCATATG 360  
DB GCATGCGCCACGCGCGCGACGCGGTTTCGCGGGAAGCAAGCTCGATATCATATG 290145

QY 361 TCGCGCGCGCTTCAATCAGCGAGTCGACGAGTTCATAATATATCTCCGACAC 420  
DB TCGCGCGCGCTTCAATCAGCGAGTCGACGAGTTCATAATATATCTCCGACAC 290085

QY 421 GGGACGAGCGCGCTTCCGCAATCTGAACTGAACTGAACTGAACTGAACTGAACTG 480  
DB GGGACGAGCGCGCTTCCGCAATCTGAACTGAACTGAACTGAACTGAACTGAACTG 290025

QY 481 CGCTCGCGGAGCGCTGCGTGGCGCAATCTTCCGCGATGTCGCGCGATTCGCAAA 540  
DB CGCTCGCGGAGCGCTGCGTGGCGCAATCTTCCGCGATGTCGCGCGATTCGCAAA 289965

QY 541 AAATCGCGCGGAACTCGACGATATCGAAGCGGAGTTTGTATTTGGTCAAGTACGCG 600  
DB AAATCGCGCGGAACTCGACGATATCGAAGCGGAGTTTGTATTTGGTCAAGTACGCG 289905

QY 601 GAAAGTACGCGCGATCGCGCGACAGCGCGGATTTGCTCGCTCGCTTCGCG 660  
DB GAAAGTACGCGCGATCGCGCGACAGCGCGGATTTGCTCGCTCGCTTCGCG 289845

QY 661 ATAACGCTGTTTGTGACCGTTCAGGCTGTGTGTCAT 695  
DB ATAACGCTGTTTGTGACCGTTCAGGCTGTGTGTCAT 289810

RESULT 13  
AAA81489/c  
ID AAA81489 standard; DNA; 837096 BP.

XX AAA81489;  
AC 04-DEC-2000 (first entry)  
XX N. meningitidis partial DNA sequence gnm\_37 SEQ ID NO:37.  
XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
XX antigen; vaccine; diagnosis; infection; antibacterial; identification;  
XX Meningococcus B; MenB; ds.  
XX Neisseria meningitidis.  
XX Neisseria meningitidis.  
XX WO200022430-A2.  
XX 20-APR-2000.  
XX 08-OCT-1999; 99WO-US23573.  
XX 09-OCT-1998; 98US-0103794.  
XX 30-APR-1999; 99US-0132068.  
XX (CHIR) CHIRON CORP.  
XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC,  
XX Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
XX Rappuoli R, Pizza M;  
XX WPI; 2000-318079/27.  
XX Isolated nucleotide sequences of Neisseria meningitidis which can be  
XX used in the diagnosis and treatment of N. meningitidis infection and  
XX other Neisseria infections, for example, N.gonorrhoea -  
XX Claim 7; Page 629-865; 1760pp; English.

CC The present invention describes methods of obtaining immunogenic  
CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414  
CC represent specifically claimed Neisseria meningitidis genomic DNA  
CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent  
CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to  
CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to  
CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF  
CC sequences, which are all used in the exemplification of the present  
CC invention. The nucleic acid sequences, protein sequences, and antibodies  
CC against them, can be used in the manufacture of a composition. The  
CC composition can be used as a medicament (or in the manufacture of a  
CC medicament) for treating, preventing or diagnosing infection due to  
CC Neisseria bacteria. For example, some of the identified proteins could  
CC be components of vaccines against Meningococcus B; against all serotypes;  
CC and/or against all pathogenic Neisseriae. Identification of sequences  
CC from the bacterium will also facilitate production of biological probes,  
CC particularly organism-specific probes. Attempts to make efficacious  
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.  
CC Multivalent vaccines have also been tried but none have successfully  
CC overcome antigenic variability. The provision of further, complete  
CC sequences may provide an opportunity to identify secreted or surface  
CC exposed proteins that may be presumed targets for the immune system and  
CC which are not antigenically variable or at least more conserved than  
CC other more variable regions.

XX Sequence 837096 BP; 207534 A; 227065 C; 205215 G; 197280 T; 2 other;

Query Match 21.9%; Score 639; DB 21; Length 837096;  
Best Local Similarity 95.0%; Pred. No. 3.9e-164;  
Matches 660; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 1 GAGTTTGGCTCCCGACGACGATGATGCTTGAGCCGACACCTGTCCGCGAAGCGG 60  
DB 55275 GAGTTTGGCTCCCGACGACGATGATGCTTGAGCCGACACCTGTCCGCGAAGCGG 55216  
QY 61 CTGACCGCCCTTTTGGCCCATCGACATCGTAACATCGTGTGGTGGCAAGCTCTTTC 120

Db 55215 CTGACCGCCCGCTCTGCCCATCGACATCGTAACAATCGGTTGCGGCAAGCTCTTTC 55156  
Qy 121 GCTTTGAGGCTGCGCAGAAAGTCAACAGTCTTCCGCGTCTTCCGCGATCACGCA 180  
Db 55155 GCTTTGAGGCTGCGGAAAGTCAACAGTCTTCCGCGTCTTCCGCGATCACGCA 55096  
Qy 181 ATTTTCAGATGTCGCGCGGAGTCTCCATCTGTTTCAGACGCGATACGATTTCTT 240  
Db 55095 ATTTTCAGATGTCGCGCGGAGTCTCCATCTGTTTCAGACGCGATACGATTTCTT 55036  
Qy 241 TCGCGCGGCTGCGGTGAACCTCAATTCAGACGAGCGCGCGATGCGGTTTTTGA 300  
Db 55035 TCGCGCGGCTGCGGTGAACCTCAATTCAGACGAGCGCGCGATGCGGTTTTTGA 54976  
Qy 301 GCATGCGCAGCGCGCGCGGAGTCTTCCCGGAAAGCTCGATATCGATAATG 360  
Db 54975 GCATTTCCACGCGCGCACCGCGGAGTCTTCCCGGAAAGCTCGATATCGATAATG 54916  
Qy 361 TCGCGCAGCGCGCTTTCATTCAGCGAGTTCGAGCAGTTCGAGTTCGAGTTCGAG 420  
Db 54915 TCGCGCAGCGCGCTTTCATTCAGCGAGTTCGAGCAGTTCGAGTTCGAGTTCGAG 54856  
Qy 421 GGAAACAGCGCGCTTTCGCGCATGCGTTCGAAAGTTCGAAAGTTCGAAAGTTCGAA 480  
Db 54855 GGAAACAGCGCGCTTTCGCGCATGCGTTCGAAAGTTCGAAAGTTCGAAAGTTCGAA 54796  
Qy 481 GCGTCGCGGAGCGTCTGCGGTGCGCAATCTTCCGCGATGTCGCGCGATTCGAA 540  
Db 54795 GCGTCGCGGAGCGTCTGCGGTGCGCAATCTTCCGCGATGTCGCGCGATTCGAA 54736  
Qy 541 AAATCGCGCGGAACTCGAGTATCGAAGGCGAGTTCGATTTGTTGTTGTTGTTGTTG 600  
Db 54735 AAATCGCGCGGAACTCGAGTATCGAAGGCGAGTTCGATTTGTTGTTGTTGTTGTTG 54676  
Qy 601 GAAATACGCGCGCATCGCGCGGAGCGAGCGCGAGTTCGATTTGTTGTTGTTGTTGTTG 660  
Db 54675 GAAATACGCGCGCATCGCGCGGAGCGAGCGCGAGTTCGATTTGTTGTTGTTGTTGTTG 54616  
Qy 661 ATAACGCGTCTTTGACGCTGCGCTGCTGTCAT 695  
Db 54615 ATAACGCGTCTTTGACGCGTTCAGGCTGCTGTCAT 54581

## RESULT 14

ABZ40085/c

ID ABZ40085 standard; DNA; 762 BP.

XX AC ABZ40085;

XX DT 07-MAR-2003 (first entry)

XX DE N. gonorrhoeae nucleotide sequence SEQ ID 4759.

XX KW Antibacterial; infection; vaccine; gene therapy; gene; ds.

XX OS Neisseria gonorrhoeae.

XX PN WO200279243-A2.

XX PD 10-OCT-2002.

XX PF 12-FEB-2002; 2002WO-IB02069.

XX PR 12-FEB-2001; 2001GB-0003424.

XX PA (CHIR-) CHIRON SPA.

XX PI Fontana MR, Pizza M, Masignani V, Monaci E;

XX DR WPI; 2003-058415/05.

XX DR P-PSDB; ABP75115.

XX XX

PT New protein from *Neisseria gonorrhoeae*, useful for the manufacture of a  
PT medicament for treating or preventing *N. gonorrhoeae* infection -  
XX  
PS Disclosure; Page 536; 815pp; English.  
XX  
CC The present invention relates to proteins from *Neisseria gonorrhoeae*.  
CC Also disclosed are the nucleic acid molecules encoding the proteins and  
CC antibodies that specifically bind to the proteins. The composition  
CC comprising the protein, nucleic acid or antibody is useful for the  
CC manufacture of a medicament for treating or preventing *N. gonorrhoeae*  
CC infection, this may be in the form of a vaccine or gene therapy.  
CC Sequences given in records ABZ37706-ABZ42016 represent nucleic acid  
CC molecules of the invention.

XX  
SQ Sequence 762 BP; 161 A; 238 C; 214 G; 149 T; 0 other;

Query Match 21.9%; Score 637.4; DB 25; Length 762;

Best Local Similarity 94.8%; Pred. No. 3.7e-165;

Matches 659; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 1 GAGTTTTCGTTCCCGAACCGAACTGATGCTTGAGCCGAAACACCTGTCGCGCAAGCGG 60

Db 695 GAGTTTTCGTTCCCGAACCGAACTGATGCTTGAGCCGAAACACCTGTCGCGCAAGCGG 636

Qy 61 CTGACCGCCCTTTTGGCCCATCGACATCGTAAACATCGGTTTGGTGGCAAGCTCTTTC 120

Db 635 CTGACCGCCCTTTTGGCCCATCGACATCGTAAACATCGGTTTGGTGGCAAGCTCTTTC 576

Qy 121 GCTTTGAGCGTGGCAGAAAGTCAAGTCAAGTCTTCCGCGTTCGCGCATCACCGCA 180

Db 575 GCTTTGAGCGTGGCAGAAAGTCAAGTCAAGTCTTCCGCGTTCGCGCATCACCGCA 516

Qy 181 ATTTTGCAGATGTCGCGCGCGAGTCTTCCATCTGTTTCAGACGCGATACGATTTCTTC 240

Db 515 ATTTTGCAGATGTCGCGCGCGAGTCTTCCATCTGTTTCAGACGCGATACGATTTCTTC 456

Qy 241 TCGCGCGCGTTCGCGTGAACCTCATGATTCAGACGCGCGCGCGATGTCGCGTTTTTGA 300

Db 455 TCGCGCGCGTTCGCGTGAACCTCATGATTCAGACGCGCGCGCGATGTCGCGTTTTTGA 396

Qy 301 GCATGCGCCACCGCGCGCGCGAGTCTTCCGCGGAAAGCTTCGATATCGATAATG 360

Db 395 GCATTTGCCACCGCGCGCGAGTCTTCCGCGGAAAGCTTCGATATCGATAATG 336

Qy 361 TCGCGCAGCGCGCTTTCATTCAGCGAGTTCGAGCAGTTCGAGTTCGAGTTCGAG 420

Db 335 TCGCGCAGCGCGCTTTCATTCAGCGAGTTCGAGCAGTTCGAGTTCGAGTTCGAG 276

Qy 421 GGGAACGAGCGCGCTTTCGCGCATGCGTTCGAAAGTTCGAAAGTTCGAAAGTTCGAA 480

Db 275 GGGAACGAGCGCGCTTTCGCGCATGCGTTCGAAAGTTCGAAAGTTCGAAAGTTCGAA 216

Qy 481 GCGTCGCGCAGCGTCTGCGTGGCGCAATCTTCCGCGATGCTGCGCGCATTCGCAAA 540

Db 215 GCGTCGCGCAGCGTCTGCGTGGCGCAATCTTCCGCGATGCTGCGCGCATTCGCAAA 156

Qy 541 AAATCGCGCGGAACTCGACGATATCGAAGGCGAGTTCGATTTGTTGTTGTTGTTG 600

Db 155 AAATCGCGCGGAACTCGACGATATCGAAGGCGAGTTCGATTTGTTGTTGTTGTTG 96

Qy 601 GAAATACGCGCGCATTCGCGCGCGCAAGCGCGCAGCGGATTTGTCGCTTCGCG 660

Db 95 GATAAATCGCGCGCATTCGCGCGCGCAAGCGCGCAGCGGATTTGTCGCTTCGCG 36

Qy 661 ATAACGCGTCTTTTTCAGCGTTCAGGCTGCTGTCAT 695

Db 35 ATAACGCGTCTTTTTCAGCGTTCAGGCTGCTGTCAT 1

## RESULT 15

AAS59546/c

ID AAS59546 standard; DNA; 29559 BP.

XX XX

AC AAS59546;  
XX 13-FEB-2002 (first entry)  
XX Propionibacterium acnes immunogenic protein encoding DNA #41.  
XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KW dermatological; osteopathic; neuroprotectant; ds.  
XX  
OS Propionibacterium acnes.  
XX  
PN WO200181581-A2.  
XX  
PD 01-NOV-2001.  
XX  
XX 20-APR-2001; 2001WO-US12865.  
XX  
XX 21-APR-2000; 2000US-199047P.  
PR 02-JUN-2000; 2000US-208841P.  
PR 07-JUL-2000; 2000US-216747P.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
PI  
XX WPI; 2001-616774/71.  
DR  
XX Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris -  
XX  
XX Claim 1; SEQ ID No 41; 1069pp; English.  
XX  
XX Sequences AAS59506-AAS59804 represent DNA molecules encoding  
CC Propionibacterium acnes immunogenic polypeptides. The proteins and their  
CC associated DNA sequences are used in the treatment, prevention and  
CC diagnosis of medical conditions caused by P. acnes. The disorders include  
CC SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and  
CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved  
CC in infections of bone, joints and the central nervous system, however it  
CC is particularly involved in the inflammatory lesions associated with acne  
CC vulgaris. A method for detecting the presence or absence of P. acnes in a  
CC patient comprises contacting a sample with a binding agent that binds to  
CC the proteins of the invention and determining the amount of bound protein  
CC in the sample. The polypeptides may be used as antigens in the production  
CC of antibodies specific for P. acnes proteins. These antibodies can be  
CC used to downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the  
CC polypeptides shown in AAU49884-AAU50191 and AAU67524-AAU67527.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 29559 BP; 5419 A; 8656 C; 9368 G; 6114 T; 2 other;  
Query Match 10.2%; Score 297.2; DB 23; Length 29559;  
Best Local Similarity 52.8%; Pred. No. 1.3e-70;  
Matches 790; Conservative 0; Mismatches 658; Indels 48; Gaps 5;  
QY 1189 AAAGCTATTCCCAACGCACTCATCTCTTAAAGATATCGATCGCGCGGCAACACC 1248  
DB 21955 AACGGTGTGCGGACGCTCTGATGATCTCAAGCGCTCGACGAGCCGCTCTCGGAGC 21896  
QY 1249 CCGATTGGATTTTGGCCAAACAAACAGTCGGCGGGGTGTGCTACGTGATTTGTTGGCG 1308  
DB 21895 CCGACTGGCTTCAACAAACGACGATGATCGTTATGCAACCTATACCGACCATTTCTCG 21836  
QY 1309 GCGATTGAAGGGCTTGAAGATAAAATTCCTTATTTTCAAGAGCTTGGTTGACTTATC 1368

DB 21835 GCACCTCAAAGGTATAAGCGATCACCTCGATCACTCTCGCATGGGTGTGCGCTATC 21776  
QY 1369 TGACCTGATGCGCGCTGTTTAAATGCCTCAAGCGAAAGCGACGCGGGCTATCGGCTCA 1428  
DB 21775 TGCATCTCATGCCCCCTGCTCAACCTCGCCAGGTACTCTGACGATGGTGTATCGGTTG 21716  
QY 1429 GCGACTACCGGATGTAATCGGCACTGGGCAATAGGAGGACTTGGCGGAAGTCAATTG 1488  
DB 21715 CGGACCATCGACCTATCCGCAOAGATCTGGGTACTATGAGATGACCTAGCTGACCTACG 21556  
QY 1489 CTGGCTGCACGAAGCGCGCATTTCCCGCTGCGATTTTATCTTCAACACCACTTCCA 1548  
DB 21555 CTACCTGCGTGGCCACGCGCATATCCCTGCTCATGACCTCATCGTTAACACCGTAGCG 21596  
QY 1549 ACGAACAGAAATGGGCGCAACGCTGCGCGCGCGGCGACCGCTTTCGACAAATTTCTACT 1608  
DB 21595 CCGAGCATGAATGGGCTCGAAGAGCTCGGGCTGGTCAACAGAGTACCGGACTATTTCC 21536  
QY 1609 ATATTTTCCCGACGCGCGGATGCCGACCAATACGACCGCACCTCGCGGAAATCTTCC 1668  
DB 21535 ATATCTGTGACTCAAGATGAGGTTGACGCTTGGGAGAGACCTTCCGAGCGTTTTTC 21476  
QY 1669 CCGACAGCACCGCGCGGCTTCTCGCAACTGGAAGACCGAC---GCTGGGTGFGACGA 1725  
DB 21475 CGGACTTCGCGCATGGCAATTTCACTGGGACGACGACTCGCAGGGGTGGTATGGCGGA 21416  
QY 1726 CCTTCAATTCCTCCATGGGACTTGAATTACAGCAACCGTGGGTATTCCGCGCATGG 1785  
DB 21415 CCTTCAACGAATTCAGTGGGATCTCACTGGGCTAAACCTTGAGTTTTCTGGAATTC 21356  
QY 1786 CGGCGGAAATGCTGTTCTTTCGCCAACTTGGCGGTTGACATCTCGCTGATGATGCGGTTG 1845  
DB 21355 TCGACCTGATGCGGTGCTGTCGCAATCGCGGGTTCGAGGTGTTCCGCTCGACGCCATCG 21296  
QY 1846 CCTTTATTTGGAACAAATGGGCAAGCTCGGAAACCTCGCGACGCGCACGCTTCA 1905  
DB 21295 CCTTCACTCGAAGAACTCGGCAACCACTGTGCAAACTTTCGGAATTTCAACACATCA 21236  
QY 1906 TCCGCGGTTCAATGCGGTTATGCGTATTCGCGCGCGCGGCTGTTCTTCAAAATCCGAAG 1965  
DB 21235 CTCAGTCAATGCGTCAAGCAATACGATCTGCGCGCGCGCGCTGTTCTTCAATGCGCGATG 21176  
QY 1966 CCAATGTCACCCCGCAACAGTGTCCAAATACATATCGGCGCAGG-----ACG 2010  
DB 21175 CCAATGTTGGCGCGCGAGTCTCACGGGATATTTCCGACGCGCGCGCGCAGTGGGGAAG 21116  
QY 2011 AATGCCAAATCGGTTACACCCCTCGCAATGCAATGTTGTGAACACACCTTCCGACGC 2070  
DB 21115 TCTCGGACATGATCTATCAACAGCTCATGTTGCACTGTGAGCGCGCTAGTACCC 21056  
QY 2071 GCGAAGTCAACCTGCTCCATCAGGCGTGAACCTACGCCCAACCTTCCGAGCATACCG 2130  
DB 21055 GCGAGCTCAGCTTCATGGAACACAGCGTTGAGTCGACGCGCGCAACACCTTCGACACAA 20996  
QY 2131 CTTGGGTCACTAGCTCGGACGACGACGACATCGGCTGGAGCTTTGCGGATGAGACG 2190  
DB 20995 CTTGGGCGACCTACGCTCGATGTCACGACGATCGGATGACCGCTGATGATGCGGACG 20936  
QY 2191 CGGCATATCTGGGCATAAAGCGGCTTACGACCAACCGCCCAATTCCTCAACGCTTCTTGGTCA 2250  
DB 20935 CCGGTAAACCTGGCTTGACCCCGCTCGCCACGCGCAGTTCTCTCGACTTCTATTCCG 20876  
QY 2251 ACCGTTTCGACGCGAGCTTGGCTGTGCGGTACCGTTCCATATACACCCAGCAGCAGCG 2310  
DB 20875 GGCATTTCCCGGATCTCTTCCCGCTGGTCTGGTCTTTCGAAGACACCCGTTACCGGCG 20816  
QY 2311 ACTGCGGTGTGAGTGTACAGCGCGGCTATGGGTGGGCTGG-----2352  
DB 20815 ACCGTGAATTTAGTGGTCTCTTAGCAGCGCTGGCGGGCTGGAAATCTTCCCTCGAGTCCG 20756  
QY 2353 -----CGCAAGACGATCCCCACGCGGTTGACCGGATCAAACTCTTGTGACAGCATTCCT 2406



```

Db      20755  ACGACCCAGCTGGTGTGAGCGTGCATCGCCGGATCGTTATGCTGCACACCGCGATCC 20696
Qy      2407  TGAGTACCGCGCGTCTGCCCGTGAATTAAGGAGGAGGAGTGGGTACGCTCAATGACG 2466
Db      20695  TCGGGTACCGCGGAGTACCCTAACTGGATGGGACGAGGTTCGGATGCTCA---ACG 20639
Qy      2467  ACGACTGGTCGCAAGACAGCAATAGAGCGACGACCGGTTGGGGCGACCGTCCGGCGCT 2526
Db      20638  ACGACTGGCAACGTGATCCCGGTCTATGCCGACGACACCGCTGGGTCCATCGACCTATGA 20579
Qy      2527  ACAACGAAGCCCTGTATCGGCGCA---ACGCAACGATCCGTGACCGCAGCCGGGCAAACT 2583
Db      20578  TGAACCTGGTCGATGGTCAAGCAGGCCACCGCGAGCCGACAGCGTTCTGGTCAATAT 20519
Qy      2584  ATCAGGGCTTGGCCATATGATGCCGTGGCCAAAGCAATCGCGCTTCGAAGGC 2639
Db      20518  GGAACGGGTGGCGCCGCGCCATCAATGCCCCGACACCGCGAGCCCCGAGTTCCATGCC 20463

```

Search completed: November 8, 2003, 13:08:18  
 Job time : 743 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 13:08:21 ; Search time 193 Seconds  
(without alignments)  
6664.199 Million cell updates/sec

Title: US-09-843-007A-1  
Perfect score: 2914  
Sequence: 1 GAGTTTGGTTCCGACACC.....CGTTTCAGCGCAATTCG 2914

Scoring table: IDENTITY\_NUC  
Gapop 10\_0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*

1: /cgn2\_6/prodata/2/ina/5A COMB seq:\*

2: /cgn2\_6/prodata/2/ina/5B COMB seq:\*

3: /cgn2\_6/prodata/2/ina/6A COMB seq:\*

4: /cgn2\_6/prodata/2/ina/6B COMB seq:\*

5: /cgn2\_6/prodata/2/ina/6C COMB seq:\*

6: /cgn2\_6/prodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2914	100.0	2914	3	US-08-737-752A-1 Sequence 1, Appli
2	125.2	4.3	3414	4	US-09-252-991A-8258 Sequence 8258, Ap
3	125.2	4.3	4125	4	US-09-252-991A-8065 Sequence 8065, Ap
4	104.4	3.6	1704	1	US-08-528-199-2 Sequence 2, Appli
5	104.4	3.6	1704	1	US-08-528-199-5 Sequence 5, Appli
6	99.4	3.4	4403765	3	US-09-103-840A-2 Sequence 2, Appli
7	99.4	3.4	4411529	3	US-09-103-840A-1 Sequence 1, Appli
8	86.6	3.0	768	4	US-09-107-532A-2141 Sequence 2141, Ap
9	84.6	2.9	873	4	US-09-328-352-3264 Sequence 3264, Ap
10	73.6	2.5	2889	1	US-08-537-002A-4 Sequence 4, Appli
11	73.6	2.5	2889	3	US-08-863-010-4 Sequence 4, Appli
12	73.6	2.5	2889	3	US-09-024-402A-4 Sequence 4, Appli
13	73.6	2.5	3600	1	US-08-537-002A-5 Sequence 5, Appli
14	73.6	2.5	3600	3	US-08-863-010-5 Sequence 5, Appli
15	73.6	2.5	3600	3	US-09-024-429-5 Sequence 5, Appli
16	59.8	2.1	1782	1	US-08-374-155A-13 Sequence 13, Appli
17	59.8	2.1	1782	2	US-08-785-396-13 Sequence 13, Appli
18	58.6	2.0	753	4	US-09-134-001C-2602 Sequence 2602, Ap
19	58	2.0	4776	2	US-08-852-401-1 Sequence 1, Appli
20	57.6	2.0	1728	4	US-09-107-532A-2079 Sequence 2079, Ap
21	49.4	1.7	1635	4	US-09-107-532A-482 Sequence 482, Appl
22	47	1.6	4403765	3	US-09-103-840A-2 Sequence 2, Appli
23	47	1.6	4411529	3	US-09-103-840A-1 Sequence 1, Appli
24	46.8	1.6	1704	1	US-08-374-155A-15 Sequence 15, Appli
25	46.8	1.6	1704	2	US-08-785-396-15 Sequence 15, Appli
26	46.4	1.6	1509	4	US-09-252-991A-11048 Sequence 11048, A
27	46.4	1.6	1590	4	US-09-252-991A-11206 Sequence 11206, A

C 28 45.8 1.6 4105 4 US-09-634-238-182 Sequence 182, Appli

29 45.8 1.6 4784 4 US-09-634-238-24 Sequence 24, Appli

30 43.4 1.5 1842 4 US-09-252-991A-5853 Sequence 5853, Ap

31 43.4 1.5 2034 4 US-09-252-991A-5871 Sequence 5871, Ap

32 43.4 1.5 6142 4 US-09-514-302-1 Sequence 1, Appli

33 43.4 1.5 1488 4 US-09-252-991A-5798 Sequence 5798, Ap

34 42.4 1.5 77536 4 US-09-410-551B-1 Sequence 1, Appli

35 41.4 1.4 1872 3 US-08-818-111-17 Sequence 17, Appli

36 41.4 1.4 1872 4 US-09-056-556-17 Sequence 17, Appli

37 41.4 1.4 1872 4 US-09-072-596-17 Sequence 17, Appli

38 41.4 1.4 1650 4 US-09-107-532A-3186 Sequence 3186, Ap

39 41 1.4 2244 1 US-08-476-519-10 Sequence 10, Appli

40 41 1.4 2244 5 PCT-US95-09323-10 Sequence 10, Appli

41 41 1.4 2334 1 US-08-476-519-1 Sequence 1, Appli

42 41 1.4 2334 5 PCT-US95-09323-1 Sequence 1, Appli

43 41 1.4 2334 5 PCT-US95-09323-1 Sequence 59, Appli

44 40.8 1.4 9223 4 US-08-961-527-59 Sequence 1, Appli

C 45 40 1.4 77536 4 US-09-410-551B-1 Sequence 1, Appli

#### ALIGNMENTS

RESULT 1  
US-08-737-752A-1  
; Sequence 1, Application US/08737752A  
; Patent No. 6265635  
; GENERAL INFORMATION:  
; APPLICANT: Rossmann, Jens  
; APPLICANT: Buttcher, Volker  
; APPLICANT: Welsh, Thomas  
; TITLE OF INVENTION: DNA SEQUENCES ENCODING ENZYMES CAPABLE  
; OF FACILITATING THE SYNTHESIS OF LINEAR  
; TITLE OF INVENTION: ALPHA-1,4 GLUCANS IN PLANTS, FUNGI AND  
; MICROORGANISMS  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: James F. Haley, Jr., c/o Fish & Neave  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States of America  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/737,752A  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P 44 17 879.4  
; FILING DATE: 18-MAY-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P 44 47 388.5  
; FILING DATE: 22-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: James F. Haley, Jr., Esq.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: GFB-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 596-9000  
; TELEFAX: (212) 596-9090  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2914 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO

Query Match	100.0%;	Score 2914;	DB 3;	Length 2914;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2914;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	GAGTTTTCGGTTCCCGAACCGAAACGTGATGCTTTCAGCCGGAACACCTGTCCGCGCAAGCGG	60	
Db	1	GAGTTTTCGGTTCCCGAACCGAAACGTGATGCTTTCAGCCGGAACACCTGTCCGCGCAAGCGG	60	
Qy	61	CTGACGCCGCCCTTTTCCCATCGACATCGTAACTTCGTTGGTGGCAAGCTCTTTC	120	
Db	61	CTGACGCCGCCCTTTTCCCATCGACATCGTAACTTCGTTGGTGGCAAGCTCTTTC	120	
Qy	121	GCITTTGAGCGTGGCAGAAAGCAAAAGTCACACGCTTTCGCCGCTTTGGCGCATCACCGCA	180	
Db	121	GCITTTGAGCGTGGCAGAAAGCAAAAGTCACACGCTTTCGCCGCTTTGGCGCATCACCGCA	180	
Qy	181	ATTTTGACAGATGTCGGGCCCGAGTCCTCCATCTGTTCAGACGCGATACGATTCTTCT	240	
Db	181	ATTTTGACAGATGTCGGGCCCGAGTCCTCCATCTGTTCAGACGCGATACGATTCTTCT	240	
Qy	241	TGCGCGCGCGTGGCGGTGAAACTCATGATTCAGACAGCGGGCGGCGATCCGCTTTTTTGA	300	
Db	241	TGCGCGCGCGTGGCGGTGAAACTCATGATTCAGACAGCGGGCGGCGATCCGCTTTTTTGA	300	
Qy	301	GCATGCGCACGGCGCGCGGCGGATGTCGCCGGAAGAAAGCTCGATATCGATAATG	360	
Db	301	GCATGCGCACGGCGCGCGGCGGATGTCGCCGGAAGAAAGCTCGATATCGATAATG	360	
Qy	361	TCGGCGAGCGCGCTTTCANTCAGCGAGTCGAGCAGTTCAAAATAATAATATCGTCCGAAAC	420	
Db	361	TCGGCGAGCGCGCTTTCANTCAGCGAGTCGAGCAGTTCAAAATAATAATATCGTCCGAAAC	420	
Qy	421	GGGAACGAGCGCGCTTCGCCATCGCGTCTGAAACGTAACAGCAGCGGCTTTCGGGCGAC	480	
Db	421	GGGAACGAGCGCGCTTCGCCATCGCGTCTGAAACGTAACAGCAGCGGCTTTCGGGCGAC	480	
Qy	481	CGCTTCGCGAAGCGCTTCGCTGTGGCGCAATCTTCGCCGATGCTGCCGCGCATTCCAA	540	
Db	481	CGCTTCGCGAAGCGCTTCGCTGTGGCGCAATCTTCGCCGATGCTGCCGCGCATTCCAA	540	
Qy	541	AAATCGGCGCGAACTTCGACGATATCGAAGGCGCAGGTTTTTGATTGGTCAAGTACGGCG	600	
Db	541	AAATCGGCGCGAACTTCGACGATATCGAAGGCGCAGGTTTTTGATTGGTCAAGTACGGCG	600	
Qy	601	GAAAGTACGGCGGCMATCGCGGCGCAACAGCGGCAACGCGATTTTGGTGGGTTCGGTCCG	660	
Db	601	GAAAGTACGGCGGCMATCGCGGCGCAACAGCGGCAACGCGATTTTGGTGGGTTCGGTCCG	660	
Qy	661	ATAAGCGGTGTTTTTCAGCGGTCAAGGCTGGTGTGTCATGCGGCTTGTTCGGGCTCAAAAGCA	720	
Db	661	ATAAGCGGTGTTTTTCAGCGGTCAAGGCTGGTGTGTCATGCGGCTTGTTCGGGCTCAAAAGCA	720	
Qy	721	GGTAAAGACGCAATATAGCAAGGCGACAGGCAATGTTTTTCAGCGGCAATTTCTGTGCGGC	780	
Db	721	GGTAAAGACGCAATATAGCAAGGCGACAGGCAATGTTTTTCAGCGGCAATTTCTGTGCGGC	780	
Qy	781	CGGCTTGATATCAATCAACGACGATCCGCGATATCGGAATCGAGCTTTCGCAAGCCCTG	840	
Db	781	CGGCTTGATATCAATCAACGACGATCCGCGATATCGGAATCGAGCTTTCGCAAGCCCTG	840	
Qy	841	TCCTTTCTAGTCAGTCGCGGAGTTCTTTCAGTATGATTCGACGCAACGCCCTTACACGGCAT	900	



Db 993 GGAGGATCGCTATCCGATCAGCAGCATCTGCGCCAGACCCCGGACATCCCGGCAATTG 1052  
Qy 2129 CGCCTGGTCACTAGCTCGGAGCAGCAGCAGCATCGCTGAGCGTTTGGCCGATGAAGA 2188  
Db 1053 CCAATGGCGGATCTTCTCGGCAACACGACGAGCTGACCTGGAGATGGTCAACCGAGA 1112  
Qy 2189 CGCGGCATATCTGGGATAGCGGCTACGACACACCGCAATTCCTCAACCGGTTCTTCT 2248  
Db 1113 CGAGCGGACTATCTCTGAGACCATCTATGCGCGCGACCGCGCGCGCTCAACCTGG 1172  
Qy 2249 CAACCG 2254  
Db 1173 CATCG 1178

RESULT 3  
US-09-252-991A-8065/c  
; Sequence 8065, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 8065  
; LENGTH: 4125  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-8065

Query Match 4.3%; Score 125.2; DB 4; Length 4125;  
Best Local Similarity 48.2%; Pred. No. 3.5e-24;  
Matches 466; Conservative 0; Mismatches 473; Indels 27; Gaps 3;

Qy 1307 CGCGGATTTCAAGGGCTTGAAGATAAAATTCCTTATTTTCAAGAGCTTGGTTTGACTTA 1366  
Db 3670 CGCGGACTTCGCGGGCTGATGAGAAGCTCGACTACATCGCGGACTGGCGGTGAACAC 3611  
Qy 1367 TCTGACCTGATCGGCTGTTTAAATGCTTGAAGGCAAGGCGGCGGCTATGGGT 1426  
Db 3610 TCTGCTGCTGCGGTT-----CTACCGGTGCGGACCGCGGCGGCTACGACAT 3557  
Qy 1427 CAGCAGCTACCGGATGTAATCGGCACTGGGCAATAGGCGACTTGGCGGAAGTCAT 1486  
Db 3556 CGCCAGTACGTTGGGTGACAGGACTAGCGGAGCTCGCGGCGGCGGCTGAT 3497  
Qy 1487 TGCTGGCTGACGAAAGCGGCAATTCGCGGCTGCTGATTTTATCTTCAACACACCTC 1546  
Db 3496 CGCGAGGCGCATCGACCGGCTGCGGGTGATTAACGAGCTGGTATCAACACACCTC 3437  
Qy 1547 CAAGACACGATGGGCGCAACGCTGCGCGCGGCGGCGGCGGCTTTTGGACAATTCTA 1606  
Db 3436 CGACGACATCCCTGGTTTATCCGCGCGCGGCGGCGGCGGAGGATCGGCGCGCGG 3377  
Qy 1607 CTATATTTTCCGACCGCGGATGCGCGCAATACGACCGGCGGCGGCGGAAATCTT 1666  
Db 3376 CTACTAGCTCTGGTCGACAGCG---ACGAGAAATACCGGCTACGCGGATCATCTTCT 3320  
Qy 1667 CCGGACGACGCGCGGCGGCTTCTGCACTGGAGAGCGGCGGCTGGGTGGGACGAC 1726  
Db 3319 CGACACGAGCGGCTTCACTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3260  
Qy 1727 CTTCAATTCCTCAATGGGACTTGAATTAAGCAACCGGCGGCTATTCGCGCAATGGC 1786  
Db 3259 CTTCTATTCACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 3200

Qy 1787 GGCGAAATGCTTCTTCTTGCACACTTGGCGGTTGACATCTGCGTATGCGTATGCGGTTGC 1846  
Db 3199 CGGGTGTGCTACTTGGCTGACATGGGCTGCGGCTGCGCTGCGCTGCGGATTC 3140  
Qy 1847 CTTTATTTGAAACAAATGGGGAAGAGTGGGAAACCTTGGCGAGGCGGCGGCGGCTCAT 1906  
Db 3139 CTACCTGTATGAAAGCGGACCGACCGAGCAGAGACCTTGGCGGAGACCCACCGGTTCT 3080  
Qy 1907 CGCGCGTTCAATGCGGTTATGCGGTTATGCGGCGGCGGCGGCTTCTTCAAAATCCGAGC 1965  
Db 3079 CAAGCGGATTCGCGGCGGAGCTGGAGCGGCACTATCCGACCGCATGCTGCTGGCGGCG 3020  
Qy 1967 CATGCTCCACCCCGGACCAAGTCTGCTCAATATATGCG-----GCAGGA 2008  
Db 3019 CAACGAGTGGCGGAGACACCCCGGCGGCTTCTGGCGGCGGAGGATGGCGGCGGCGG 2960  
Qy 2009 CGATGCGCAATCGGTTTACAAACCCCGCTCAATGCAATGTTGTGAAACACCTTGGCCAC 2068  
Db 2959 CGATGCGCAATGCGCTTCCACTTCCCGCTGATGCGGCGGCACTGATGCGGCGGCGG 2900  
Qy 2069 GCGGAAATCAACCTGCTTCCATAGGCGCTGACCTACCGGCGGCAACCTGCGCGGAGCAT 2128  
Db 2899 GGAGGATCGTATCGATCAGCATCTGCGGCGGAGACCCCGGACATCCCGGCGGCAATTG 2840  
Qy 2129 CGCCTGGTCACTAGCTGCGGCGGCGGCGGCGGCGGCGGCTTGGCGGATGAAGA 2188  
Db 2839 CCAATGGGCGGATCTTCTTGGCGGCAACCGAGAGCTGACCTGGAGATGGTACCGGCGA 2780  
Qy 2189 CGCGGATATCTGGGCGGCTAAGCGGCTACGACCGGCGGCAATTCCTCAACCGGCTTCTT 2248  
Db 2779 CGAGCGGCGACTATCTCTGGAACCACTATGCGGCGGCGGCGGCGGCGGCTCAACCTGG 2720  
Qy 2249 CAACCG 2254  
Db 2719 CATCG 2714

RESULT 4  
US-08-528-199-2  
; Sequence 2, Application US/08528199  
; Patent No. 5763228  
; GENERAL INFORMATION:  
; APPLICANT: KUBOTA, Michio  
; APPLICANT: TSUSAKI, Keiji  
; APPLICANT: SUGIMOTO, Toshiyuki  
; TITLE OF INVENTION: RECOMBINANT ENZYME FOR CONVERTING  
; TITLE OF INVENTION: MALTOSE INTO TREHALOSE  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/528,199  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,126  
; FILING DATE: 14-SEP-1995  
; APPLICATION NUMBER: US 08/485,126  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 156399/1994  
; FILING DATE: 16-JUN-1994  
; ATTORNEY/AGENT INFORMATION:

NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: KUBOTA-6A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1704 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-528-199-2

Query Match 3.6%; Score 104.4; DB 1; Length 1704;  
Best Local Similarity 49.4%; Pred. No. 1.1e-18;  
Matches 304; Conservative 0; Mismatches 306; Indels 6; Gaps 1;

QY 1408 GCGACGGCGGTATGGCGGTACGAGCTACCGGATGTCATTCGGCACTGGGCAATAG 1467  
DB |||||  
QY 191 GCGACGGCGGTACGAGCTACCGGATGTCATTCGGCACTGGGCAATAG 250  
DB |||||  
QY 1468 GCGACTTGGCGGAAGTCAATTCGCTGCGTACGAGCGCGGATTCGCGCGTCTGCGATT 1527  
DB |||||  
QY 251 AGGACTTCCAGCCCTTCTCTGACGCGCGGCGACGAGCGCGGATCCGGTGCATCATCGACT 310  
DB |||||  
QY 1528 TTATCTTCAACCACTCCCAACGAAACACGAAATGGGCGCAACGCTGCGCGCGCGGACC 1587  
DB |||||  
QY 311 TCGTATGAACACACGAGTGACGCGCGCGGCTTCAGGCTTCGCGAGGATCCG 370  
DB |||||  
QY 1588 CGCTTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGCGGATTCGCGCAATAGGACC 1647  
DB |||||  
QY 371 ACGGCCCGTACGGCGACTTCTACGCTCTGGTCCGACACGAGCTCTACCGAGCG--- 427  
DB |||||  
QY 1648 GCACCTTCGCGGAATCTTCCCGACCGACGACCGCGGCTTCTCGCACTGGAGACG 1707  
DB |||||  
QY 428 ---CGCGGTGATCTTCTGTCGACCGGCGGATGCGTCTGCGTTCGACATGGGCTTCGACGCT 484  
DB |||||  
QY 1708 GACGCTGGGTGTGACGACGCTTCAATTCCTTCCAAATGGGACTTCAATTACAGCAACCGGT 1767  
DB |||||  
QY 485 GCGAGTACTTGGCAACCGCTTCTTCCACCAACGCGCGGACCTGAATTCGACCAACCGCA 544  
DB |||||  
QY 1768 GCGTATTCGCGCAATTCGCGCGGCGGAAATGCTGTCTTCCCACTTGGCGGTGACATCC 1827  
DB |||||  
QY 545 AGTCCAGGACCGCATGCTGAGGCGGATGGGCTTCTGGCTTCGACATGGGCTTCGACGCT 604  
DB |||||  
QY 1828 TCGGTATGAGTGGGTGCGCTTTATTTTGAACAAATGGGCAACAGCTGCGGAAACCTGC 1887  
DB |||||  
QY 605 TCGGCTGACGCGGTGCGCTTACCTCTACGAGCGTCCCGGCAACCAACGCGGAGAACCTCC 664  
DB |||||  
QY 1888 GCGAGGCGACGCTTCATCCGCGGCTTCAATGCGGTTATGCTTTCGCGCGCGCGCG 1947  
DB |||||  
QY 665 CCGAGACGACGAGATGCTCAAGCGGCTGCGGCGCTTCGTCGACGACATACCCCGGACC 724  
DB |||||  
QY 1948 TGTCTTCAATTCGGAAGCATCGTCCACCGCGGACCAATCGTCCAAATACATCGGCGGAG 2007  
DB |||||  
QY 725 GGGTCTGCTGTACGAGGCGGACCAAGTGGCGCGGACCGGACGCTGTGTGAGTACTTTCGCGCGCG 784  
DB |||||  
QY 2008 ACGAATGCCAATCGG 2023  
DB |||||  
QY 785 AGGAGCGTGAGACCG 800  
DB |||||

## RESULT 5

US-08-528-199-5  
Sequence 5, Application US/08528199  
Patent No. 5763228  
GENERAL INFORMATION:  
APPLICANT: KUBOTA, Michio  
APPLICANT: TSUSAKI, Keiji  
APPLICANT: SUGIMOTO, Toshiyuki

TITLE OF INVENTION: RECOMBINANT ENZYME FOR CONVERTING  
TITLE OF INVENTION: MALTOSE INTO TREHALOSE  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/528,199  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,126  
FILING DATE: 14-SEP-1995  
APPLICATION NUMBER: US 08/485,126  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 156399/1994  
FILING DATE: 16-JUN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: KUBOTA-6A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1704 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1704  
US-08-528-199-5

Query Match 3.6%; Score 104.4; DB 1; Length 1704;  
Best Local Similarity 49.4%; Pred. No. 1.1e-18;  
Matches 304; Conservative 0; Mismatches 306; Indels 6; Gaps 1;

QY 1408 GCGACGGCGGTATGGCGGTACGAGCTACCGGATGTCATTCGGCACTGGGCAATAG 1467  
DB |||||  
QY 191 GCGACGGCGGTACGAGCTACCGGATGTCATTCGGCACTGGGCAATAG 250  
DB |||||  
QY 1468 GCGACTTGGCGGAAGTCAATTCGCTGCGTACGAGCGCGGATTCGCGCGTCTGCGATT 1527  
DB |||||  
QY 251 AGGACTTCCAGCCCTTCTCTGACGCGCGGCGACGAGCGCGGATCCGGTGCATCATCGACT 310  
DB |||||  
QY 1528 TTATCTTCAACCACTCCCAACGAAACACGAAATGGGCGCAACGCTGCGCGCGCGGACC 1587  
DB |||||  
QY 311 TCGTATGAACACACGAGTGACGCGCGCGGCTTCAGGCTTCGCGAGGATCCG 370  
DB |||||  
QY 1588 CGCTTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGCGGATTCGCGCAATAGGACC 1647  
DB |||||  
QY 371 ACGGCCCGTACGGCGACTTCTACGCTCTGGTCCGACACGAGCTCTACCGAGCG--- 427  
DB |||||  
QY 1648 GCACCTTCGCGGAATCTTCCCGACCGACGACCGCGGCTTCTCGCACTGGAGACG 1707  
DB |||||  
QY 428 ---CGCGGTGATCTTCTGTCGACCGGCGGATGCGTCTGCGTTCGACATGGGCTTCGACGCT 484  
DB |||||  
QY 1708 GACGCTGGGTGTGACGACGCTTCAATTCCTTCCAAATGGGACTTCAATTACAGCAACCGGT 1767  
DB |||||

D5 485 GCAGTACTACTGGACCGCTTCTTCCACACAGCCGCGACCTGAACCTTGCACACCCG 544  
QY 1768 GGGTATTCCCGCAATGGCGGGGAAATGCTGTCTTCCCAATTTGGGCTTGACATCC 1827  
D5 545 AGGTCCAGGACGCATCTGGAGCGATGCGTCTTGGCTCGACATGGGCTCGACGGCT 604  
QY 1828 TGGTATGGATGGCTTGGCTTTATTTGGAAACAAATGGGGCAAGCTGCGAAACCTGC 1887  
D5 605 TCGGCTCGACGGGCTCCCTACCTCTACAGAGGTCCTCCGCAACCAAGCGGAGAACCTCC 664  
QY 1888 CGCAGGCGCACGCCCTCATCCGCGCTTCAATGCGCTTATGCGTATTCGCGCGCCGCG 1947  
D5 665 CCAGAGCGCACAGATGCTCAAGGGGTGCGGCTTCGTCGACGACAATCCCGGACC 724  
QY 1948 TGTCTTCAATCGAAGCATGCTCCACCCGACCAAGTCGTCCTCAATATACATCGGAGG 2007  
D5 725 GGGTGTGCTGTACGAGGCGAACCAGTGGCGGACCGACGCTGGTGGAGTACTTCGGGCGCG 784  
QY 2008 ACGAATGCCAAATCGG 2023  
D5 785 AGGAGCTGAGGACGG 800

## RESULT 6

US-09-103-840A-2  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 3.4%; Score 99.4; DB 3; Length 4403765;  
Best Local Similarity 49.2%; Pred. No. 9.9e-16;  
Matches 293; Conservative 0; Mismatches 296; Indels 6; Gaps 1;  
QY 1408 GCGACGGCGGTATGCGGTGACGAGTACCGGATGTCATTCGGCACTGGGCAACAATAG 1467  
D5 152780 GCGACGGCGGTACGACATTCGGGACTTCTACAGTGTCTCCCGAATTCGGCACCGTCG 152839  
QY 1468 GCGACTTGGCGGAAGTCAATTGCTGCGCTGCACGAAAGCCGCAATTCGCCGCTGCTGAT 1527  
D5 152840 ACGAATTCGTCGCCCTGGTCGACGCGCTCACCGCGAGGTATCCGCATCATCACCAGCC 152899  
QY 1528 TTATCTTCAACACCTCCCAACGAAACGAAATGGGGCAAGCTGCGCGCGCGGAC 1587  
D5 152900 TGGTATGATATCAACATTCGAGTTCGACCCCTGGTTTCAGAGTTCGCGCGGACCCAG 152959  
QY 1588 OGCTTTTCGACAAFTTCTACTATATTTTCCCGACCGCGGATGCGCAATACGAC 1647  
D5 152960 ACGACCGTACGGTGAATTAATGCTGTGGAGCGACACGAGCGCTACACCGACGCC 153019  
QY 1648 GCACCTCGCGGAATCTTCCCGACGACGACCGCGGCTTCTCGCACTGGAAGCG 1707  
D5 153020 GGATCATCTTCG-----TCGACACCGAAGTGAATGCTGCTTCCCAACTTGGCGCTGACATCC 153073

QY 1708 GACGCTGGGTGTGAGGACCTTCAATTCCTTCAATGGGACTTGAATTACAGCAACCGGT 1767  
D5 153074 GACATTTCTACTGGCACCGAATTTCTTCTCCACCAACCGAATGAACTACGCAACCCCG 153133  
QY 1768 GGGTATTCCCGCAATGGCGGGGAAATGCTGTCTTCCCAATTTGGGCTTGACATCC 1827  
D5 153134 CCGTCAAGAGCGGATGATCGAGTCACTCGCTTTTGGCTGGCTTGGCATCGACGGT 153193  
QY 1828 TCGTATGATGCGGTGCTTTATTTGGAAACAAATGGGGCAAGCTGCGAAACCTGC 1887  
D5 153194 TTCGTTGAGACCGGTGCGCTTATCTCTTTGAAACGTTGAGGGCCCAACTGCGGAAACCTGC 153253  
QY 1888 GCGAGGCGCACGCCCTCATCCGCGCTTCAATGCGCTTATGCGTATTCGCGCGCCGCG 1947  
D5 153254 CCGAAACACACGCTTTCTCAAGCGAGTCCGCAAGTGGTGGACGAAATTCGCCGCG 153313  
QY 1948 TGTCTTCAATTCGAAAGCATGCTCCACCCGACCAAGTCTGTCATATACATCGG 2002  
D5 153314 GGGTGTGCTGACCGAAGCAATCAGTGGCGCGGAGTGTGCTGCAATATTTCCG 153368

## RESULT 7

US-09-103-840A-1  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4411529  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Query Match 3.4%; Score 99.4; DB 3; Length 4411529;  
Best Local Similarity 49.2%; Pred. No. 9.9e-16;  
Matches 293; Conservative 0; Mismatches 296; Indels 6; Gaps 1;  
QY 1408 GCGACGGCGCTATCGGTGACGAGTACCGGATGTCATTCGGCACTGGGCAACAATAG 1467  
D5 152611 GCGACGGCGGTACGACATTCGGGACTTCTACAGTGTCTCCCGAATTCGGCACCGTCG 152670  
QY 1468 GCGACTTGGCGGAAGTCAATTGCTGCGCTGCACGAAAGCCGCAATTCGCCGCTGCTGAT 1527  
D5 152671 ACGAATTCGTCGCCCTGGTCGACGCGCTCACCGCGAGGTATTCGCATCATCACCAGCC 152730  
QY 1528 TTATCTTCAACACCTCCCAACGAAACGAAATGGGGCAAGCTGCGCGCGCGGAC 1587  
D5 152731 TGGTATGATATCAACATTCGAGTTCGACCCCTGGTTTCAGAGTTCGCGCGGACCCAG 152790  
QY 1588 CGCTTTTCGACAAFTTCTACTATATTTTCCCGACCGCGGATGCGCAATACGAC 1647  
D5 152791 ACGACCGTACGGTGAATTAATGCTGTGGAGCGACACGAGCGGCTACACCGACGCC 152850  
QY 1648 GCACCTCGCGGAATCTTCCCGACCGACCGCGGCGCTTCTCGCACTGGAAGCG 1707  
D5 152851 GGATCATCTTCG-----TCGACACCGAAGTGAATGCTGCTTCCCGCTTCTCGGCC 152904  
QY 1708 GACGCTGGGTGTGAGGACCTTCAATTCCTTCAATGGGACTTGAATTACAGCAACCGGT 1767  
D5 152905 GACAGTTCTACTGGCACCGAATTTCTTCTCCACCAACCGGATCTGAACTAGCAACCCG 152964  
QY 1768 GGGTATTCGCGCAATGGCGGGCGAAATGCTGCTTCCCAACTTGGCGCTGACATCC 1827



Db 152965 CCGTGAAGAGCGGATGATGACGTATCCCGCTTTGGCTCGGCTGCGCATCGACGGGT 153024  
Qy 1828 TCGGTATGATGCGGCTTCCCTTTATTTGAAACAAATGGGGACAAGCTGCGAAAACCTGC 1887  
Db 153025 TCGGTGGAGCGGCTGCGCTATCTCTTTGAAGTGAGGCGACCACTGCGAGAACTGC 153084  
Qy 1888 CGCAGGCGCAGCGCTCATCGCGGCTTCAATGCGGTTATGCGGTATGCGCGCGCGCGCG 1947  
Db 153085 CGGAACACACAGCTTTTCTCAAGCGAGTCCGCAAGGTGGTGGAGCGGAATTCCCGCGCC 153144  
Qy 1948 TGTCTTCAATCCGAAGCCATGTCACACCGCGACCAAGTGTCCCAATACATCGG 2002  
Db 153145 GGGTGTCTAGCGAGACCAATCATGTGGCGCGGATGTCTCGATATTTCCG 153199

## RESULT 8

US-09-107-532A-2141/c  
; Sequence 2141, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 2141:

SEQUENCE CHARACTERISTICS:

LENGTH: 768 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...768

SEQUENCE DESCRIPTION: SEQ ID NO: 2141:

US-09-107-532A-2141

Query Match 3.0%; Score 86.6; DB 4; Length 768;

Best Local Similarity 45.8%; Pred. No. 5,9e-14;

Matches 299; Conservative 0; Mismatches 354; Indels 0; Gaps 0;

Qy 11 TTCCGAACCGAAGCGTATGATGCTTTGAGCCGAACACCTGTCCGGCAAGCGGCTCACCGCCC 70  
Db 691 TTAAGGAACCAAACGTCATCGAGGATCTGTCTGTTCCGCAGCCAGACGAGAAATTTTC 632  
Qy 71 CTTTTTCCCGCATCGACATCGTAACATCGTTTGGTGGCAAGCTCTTTCCGCTTTGAGCG 130  
Db 631 CCAGATCGCCCATAGACATCGTAATGATAGGCATGAGAGACAAAGCAATTTGCCGCTTAG 572  
Qy 131 TGGCAGAAACCAAGTCAGCAAGTCTTTCCGCGCTTTGCGGCATCACCGCAATTTTGAGA 190  
Db 571 TCAACTCCATCAACCGCAAAACATCTCGTTCAATTTATTTGGGAGTCACCTGCCAATTTCCGA 512  
Qy 191 TGTCCGCGCGGAGTCTCTCATCTGTTTCAGACGGCATAGATTTCTTTTGGGCGGCG 250  
Db 511 TATCAGCGCCCAAAATGTTCCATGACATTCAGTCGTTAAAGCAAGGAGGTCAGCTGGTG 452  
Qy 251 TGGCGTGAAACTCATGATTTGCAGACGAGGCGCGGATGCCGTTTTTTTGGCATGCGCCA 310  
Db 451 TTTCTTTAAATCATGACTACTGATAATTAAGGAATGTTTAAAGACTTTCATTTCTCTA 392  
Qy 311 CGGCGCGCGGACGGCGGTTTCCCGGAAAAGCTCGATATCGATATGATGTCGGGAGGC 370  
Db 391 TCAGTCCCTTTTCTAGACATTTCTACTTTGAACAACTCGATATCGACCATATCTACTAATC 332  
Qy 371 GGGTTTCAATCAGCGAGTCGAGCAGTTCAAAATAAATCGTCGGAACACGGGAACGAGC 430  
Db 331 CTGTGTATCATCATGTTTCATAGGTCGGGATCTGACGATCGAAATGTCCTTTGTTCTC 272  
Qy 431 CGCTTTCGCATGCCGCTCTGAACGTAAACAGCAGCGGCTTGTCCGGCAGCGCGTCGCGGA 490  
Db 271 CGCTTTCAGCTAAGGTACGAAACGTAAATAAATACCGCAGTCCTCAAGCATCTCATGGA 212  
Qy 491 CGGCTCGGTGTGGCGCAATCTTCCCGATGCTGCCCGGCATTCGCAAAATCGGCGC 550  
Db 211 TAAGAACCAATGTTCTGCAACTTCATCAAGAAAACCTTACGATAATGATCGACTC 152  
Qy 551 GGAACCTCGACGATATCGAAGGGCGAGGTTTTTTGATTTGGTCAAGTACGGCGGAAGTACGG 610  
Db 151 GCCATCTACTAAGTCGCGCCCTGCTTCAGTCAAGAACTGCTTCAAGTAATCTTCT 92  
Qy 611 CGGCATCGCGGCGACAGCGGCAACGGGATTTTGGTGGCTCGCTTCCGATTA 663  
Db 91 CTCGGAGCGCTCCGATAAGAGCGCGCAAAATCTTTGGTCTACCGCTTCTCTATA 39

## RESULT 9

US-09-328-352-3264/c

; Sequence 3264, Application US/09328352

; Patent No. 6562958

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; FILE REFERENCE: GTC99-03PA

; CURRENT APPLICATION NUMBER: US/09/328,352

; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 8252

; SEQ ID NO 3264

; LENGTH: 873

; TYPE: DNA

; ORGANISM: Acinetobacter baumannii

US-09-328-352-3264

Query Match 2.9%; Score 84.6; DB 4; Length 873;

Best Local Similarity 47.8%; Pred. No. 2.2e-13;

Matches 246; Conservative 0; Mismatches 269; Indels 0; Gaps 0;

Qy 52 GCAAGGCGGCTGACCGCCCTTTTCCGCCATCGATCGTACAACTCGGTTGGTGGA 111

Db 752 GCACACGAGAAATCGTACCTAAATGCCCCATCGACATTCATATATGCTTAGTGT 693

Qy 112 AGCTCTTTTCCTTTGAGCGGTGGCAGAAAGTCAAGCAGCTCTTCGCGCTTTGCGGC 171

Db 692 TCGTGGCTTACTTTAAAGTTGGTTTCATTAAAGTGAAGACATCTGTATTAGATTAGGC 633  
QY 172 ATACCGCAATTTGAGATGTCGGCGCGAGTCTCTCACTGTTTCAGACGGCATAGC 231  
Db 632 ATACAGCAATTTTGAAGACATCTGGCCCACTTGTATCTTGTTTTAAACAAGCGTTTTC 573  
QY 232 ATTCTCTCTGGCGGGCGTGGGTGAACACTCATATTGACAGCAGCGCGCGATGCG 291  
Db 572 ATTTCATCTTGGCTTGGTGTGTTTTTGGAGTCTGTATTAGACATGACAAAGCACTTTC 513  
QY 292 TTTTGTGAGCATGCGCCAGCGCGCGCGAGTGTTCGCCGAAAGTTCGATA 351  
Db 512 TTTTGTGATGCTGTGTGAACAATTTCTGAACACTACTTTTGTATCAGGAAATTTCAACA 453  
QY 352 TCGATAAGTCTGGCGAGCGGCTTTCATCAACGAGTCGAGCAGTTCAAATAAATAGC 411  
Db 452 TCTAACCGATCCATGAACGGATTTTTTAAATAGCTGATAAGTCTTCCATTAATCAGA 393  
QY 412 TCGGAACACGGGAACGAGCGCGCTTCGCCATGCGCTCTGAACGTAAACAGCAGCGGCTTG 471  
Db 392 TCACTAAATTTCAAGCTGACCCCTTCGTTTTTGTAGCAATAGTGCATCATTTGGCTTA 333  
QY 472 TCGGCGAGCGCTCGGCGAGCGTCTGCGTGTGGCGCAATCTTCGCCGATGTCGCCGG 531  
Db 332 TTGCGCTAAATCTTTTCAACTCATGCGCCCAAGGCAATCACTGTTTGTATCGCTGCA 273  
QY 532 CATTCCAAAATTCGCGCGGGAACCTCGACGATC 566  
Db 272 AGCTTAATTAATCGATACGGAACCTGCTAAATC 238

## RESULT 10

US-08-537-002A-4

Sequence 4, Application US/08537002A

Patent No. 5773282

GENERAL INFORMATION:

APPLICANT: TSUSAKI, Keiji

APPLICANT: KUBOTA, Michio

APPLICANT: SUGIMOTO, Toshiyuki

TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR

TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

ADDRESSEE: BROWDY AND NEIMARK

STREET: 419 Seventh Street, N.W., Suite 300

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/537,002A

FILING DATE: 29-SEP-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 260984/1994

FILING DATE: 01-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP No. 5773282 yet received

FILING DATE: 08-SEP-1995

ATTORNEY/AGENT INFORMATION:

NAME: BROWDY, Roger L.

REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER: TSUSAKI-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-628-5197

TELEFAX: 202-737-3528

TELEX: 248633

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:  
LENGTH: 2889 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-537-002A-4

Query Match 2.58; Score 73.6; DB 1; Length 2889;

Best Local Similarity 47.08; Pred No. 3.9e-10;

Matches 448; Conservative 0; Mismatches 479; Indels 27; Gaps 6;

QY 1307 CGGCGATTTGAAGGGCTTGAAGATAAAATTCCTTATTTTCAAGAGTGTGTTGACTTA 1366  
Db 81 CGGGGACTTTGAGGCGCTGAGCGGAAGCTTCCCTACTCTGGAGAGTCTGGGGTCAACAC 140  
QY 1367 TCTGCACCTGATGCGGTGTTTAAATGCCCTGAAGGCAAAAGCGACGCGCTATGCGT 1426  
Db 141 CCTCTGGCTCAATGCCCTTCTTCCAGTCCCC-----CTTGAGGACGACGGGTACGAT 194  
QY 1427 CAGCAGCTTACCGGATGTCAATCCGGCACTGGGCACAATAGCGACATTGCGCGAAGTCAT 1486  
Db 195 CTCGAGCTACTACAGATCTCCCGCTCCACGGACCTGAGGACTTCACCG-----T 248  
QY 1487 TGCTGGCTGCAAGAGCGGCAATTTCGCCGTCTGCGATTTTATCTTCAACACACATCT 1546  
Db 249 GAGCGAGGCCACCGCGCGGGGATGAAGGTGATCATTTAGCTCTCTGAAACCAACATCT 308  
QY 1547 CAACGACACGAATGGGCGCAACGCTGCGCGCGCGGACCCCGCTTTTCGACAAATTTCTA 1606  
Db 309 CATTGACCACTTGGTTCCAG-----GAGCGGAGGAAGCCGAATAGCCCCATGCGGA 362  
QY 1607 CTATATTTTCCCGACCGCGGATGCGGACCAATAGACGCGACCTTGCAGGAATCTT 1666  
Db 363 CT---GGTACCTGTGAGCGGACACACCGCGGAGATACAAAGGGGTCCGGGTCTCTTCAA 419  
QY 1667 CCCGACCGACACCGCGCGGCTTCTCGCACTGCGAGACGACGCTGGGTGTGAGACAC 1726  
Db 420 GGACTTTGAACTTCCAACTGGACCTTTGACCCCGTGGCAAGGCTTACTTGGACCG 479  
QY 1727 CTTCAATTTCTTCCAAATGGGACTTGAATTAAGCAACCCGCTGGGTATTTCCGCGCAATGGC 1786  
Db 480 CTTCTACTGGCAACGACCGGACCTCACTGGGACAGCCCCGAGGTGGAGAGGCGCATCCA 539  
QY 1787 GGGGCAATGCTTTCCTTGGCAACTGGGGTGTGACATCTCGGTATGATGGGTTC 1846  
Db 540 CAGGTCAATGTTCTTCTGGCGGACCTGGGGGTGACGCTTCCGCTTGGAGCGCATCCC 599  
QY 1847 CTTTATTTGGAAACAAATGGGGACAGCTGCAAAACCTGCGGAGGCGC---AGCGCT 1903  
Db 600 CTACCTCTACGAGCGGGAGGGGACCTCTCTGAGAGAACCTCCCGAGACCAATTGAGGCGGT 659  
QY 1904 CATCGCGGCTTCAATGCGGTATGCGTATTTGCGCGCGCGCGCTGTGTTCTTCAATCCGA 1963  
Db 660 GAAGCGCTGAGGAGGCGCTTGGAGGAGCGCTAGCGGCCCGGGAGAGATCTCTCTCCGCGA 719  
QY 1964 AGCATCTGTCACCCCGACCAAGTCTCCAAATACATCGGGCA---GGACGAATGCCAAAT 2020  
Db 720 GGCCAAATGTGGCGGAGGAGACCTCTCTTCTGGGACGCGGAGCGGGTCCACAT 779  
QY 2021 CGTTTAAACCCCTGCAAAATGGCAATTTGTTGTGAACACCTTTCGCCACGCGGGAAGTCAA 2080  
Db 780 GGCCTACAACTTCCCTCTGATGCCCGGATTTTATGSCCTTAAGGCGGGAGGACCGGGG 839  
QY 2081 CCTGCTCCATCAGGCGCTGACCTACCGCACAACTGCGCGAGCATACCGCTGGGTCAA 2140  
Db 840 TCCCATTTGAACCATGCTCAGGAGGGGAGGGGATCCCCGAAACCGCCAGTGGGCT 899  
QY 2141 CTAGCTCCGAGCCACGACGATCGGCTGAGCTTTCGATGAGAGCGGGCATATCT 2200  
Db 900 CTTCTCCGCAACCAACGACGAGCTCACCTTGAGAGAGGTACGAGGAGGAGCGGGAGTT 959  
QY 2201 GGGCATAGCGCTACGACACCGCCCAATTCCTCAACCGCTTCTTCGTCACACCG 2254

Db 960 CATGTACAGGCTACGCCCCCGAGCCCAAGTTCCGATCACTGGGATCCG 1013

## RESULT 11

US-08-863-010-4  
; Sequence 4, Application US/08863010  
; Patent No. 6067146  
; GENERAL INFORMATION:  
; APPLICANT: TSUSAKI, Keiji  
; APPLICANT: KUBOTA, Michio  
; APPLICANT: SUGIMOTO, Toshiyuki  
; TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR  
; TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/863,010  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/537,002  
; FILING DATE: 29-SEP-1995  
; APPLICATION NUMBER: JP 260984/1994  
; FILING DATE: 01-OCT-1994

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP No. 6087146 yet received

FILING DATE: 08-SEP-1995

ATTORNEY/AGENT INFORMATION:  
; NAME: BROWDY, Roger L.

REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER: TSUSAKI=1

TELEPHONE: 202-628-5197

TELEFAX: 202-737-3528

TELEX: 248633

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 2889 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-863-010-4

Query Match 2.5%; Score 73.6; DB 3; Length 2889;  
Best Local Similarity 47.0%; Pred. No. 3.9e-10;  
Matches 448; Conservative 0; Mismatch 479; Indels 27; Gaps 6;

QY 1307 CGCGCATTTGAAGGGCTTGAAGATATAATTCCTTATTTTCAAGAGCTTGGTTGACTTA 1366

Db 81 CGGGGACITTTGAGGCTTGAGCGGAGCTTCCCTACCTGGAGGAGCTCGGGTCAACAC 140

QY 1367 TCTGCACCTGATGCGCTTTAAATGCGCTGAAGGCAAAAGCGAGCGGGCTATGCGGT 1426

Db 141 CCTCTGGCTCATGCCCTTCTCCAGTCCCC-----CTTGGGGAGCGAGGATACGATAT 194

QY 1427 CAGCAGCTACCGGATGTCAATCCGCGACTGGGCAATAGCGGACTTGGCGGAAGTCAT 1486

Db 195 CTCGGACTACTACGATCTCCCGTCCACGGGACCTGGAGGACTTCACCG-----T 248

QY 1487 TCGTGGCTGCACGAGCGGCAATTTCCGCCGCTGTCGATTTTATCTTCAACACACACCTC 1546

Db 249 GGACGAGGCCACGCGCGGGGATGAGGTGATCATTTGAGCTGCTGTAACACACCTC 308

QY 1547 CAACGAACAGAAATGGCGCAACGCTGCGCGCGGGGACCCGCTTTTTCGACAAATTTCTA 1606

Db 309 CATTGACCACTCTGGTTCCAG-----GAGGCGAGGAAGCCGAATAGCCCCATGCGGA 362

QY 1607 CTATATTTTCCCGACCGCGGATGCCGACCAATACGACCGCACCTGCGCGGAATCTT 1666

Db 363 CT---GGTAGCTGTGGAGCGACACCCCGAGAGTACAGGGGTCGGGTCTCTTCAA 419

QY 1667 CCCGACCAACACCGCGGCTTTCGCAATGGAAGAGCGAGCTGGGTGTGGAGAC 1726

Db 420 GGACTTTGAAACCTTCAACTGGACCTTTGACCCCGCTGCGCAAGSCCTACTACTGGCACG 479

QY 1727 CTTCAATTCCTCCANTGGCACTTGATTAACAGCAACCCGCTGGGTATTCGCGCAATGGC 1786

Db 480 CTTCTACTGGACACCGCCGACCTCACTGGGACAGCCCGAGGTGGAGAGGCCATCCA 539

QY 1787 GGGCGAAATGCTGTTCTTCCCACTTGGGCTTGACATCTGCTGTATGATGCGGTGC 1846

Db 540 CCAGGTCAATGTTCTTCTGGGCGGACCTGGGCTTCCGCTTGGACGCGCATCCC 599

QY 1847 CTTTATTTGAAACAAATGGGCAAGCTGGAACCTGCGAGGCGC---ACGCCCT 1903

Db 600 CTACTCTACGAGCGGAGGAGGACCTCTCGGAGAACCTCCCGAGACCATTTGAGGCGT 659

QY 1904 CATCCGGCGCTTCAATGCCGTTATGCTATTGGCGCGCCCGCTGTTCCTTCAAAATCGA 1963

Db 660 GAACGCTCTGAGGAAGCCCTGGAGGAGGCTACGCGCCCGGGAAGATCTCTCTCGCGA 719

QY 1964 AGCATGCTCCACCCCGACCAAGCTGCTCCATATACATCGGCA---GGACGAATGCCAAT 2020

Db 720 GGGCAACATGTGGCGGAGGAGACCTCTCGGAGAACCTCCCGAGACCGGGGTCCACAT 779

QY 2021 CGGTTACAACCCCTCGAAATGSCATTTGTGGAACACACCTTGCACGCGCGAAGTCAA 2080

Db 780 GGCCTACACTTCCCTCTGATGCGCGATCTTCATGGCTTTCATGCGGAGGAGCGGGG 839

QY 2081 CCGCTCCATCAGGCGCTGACCTTACCGCCCAACCTGCGCGAGCATACCGCTGGGTCAA 2140

Db 840 TCCCATTTGAAACCATGCTCAAGGAGGCGGAGGATCCCGAAACCGCCAGTGGGCT 899

QY 2141 CTAGTCCGAGCGCAACGACGATCGGTGAGCTTTGCGATGAAGACGCGCATATCT 2200

Db 900 CTTCTCCGCAACCAACGACGAGCTACCTGGAGAGTACGCGAGGAGCGGAGTT 959

QY 2201 GGGCATAAGCGGCTACGACCAACCGCCCAATTCCTCAACCGCTTCTTCGTCAACCG 2254

Db 960 CATGTACGAGGCTTACGCCCCCGACCCCAAGTTCCGATCAACCTGGGATCCG 1013

## RESULT 12

US-09-024-429-4

; Sequence 4, Application US/09024429

; Patent No. 6165768

; GENERAL INFORMATION:

; APPLICANT: TSUSAKI, Keiji

; APPLICANT: KUBOTA, Michio

; APPLICANT: SUGIMOTO, Toshiyuki

; TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR

; TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE

; NUMBER OF SEQUENCES: 17

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BROWDY AND NEIMARK

; STREET: 419 Seventh Street, N.W., Suite 300

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20004

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
CURRENT APPLICATION DATA: Patent in Release #1.0, Version #1.30  
APPLICATION NUMBER: US/09/024,429  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/537,002  
FILING DATE: 29-SEP-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 260984/1994  
FILING DATE: 01-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 255829/1995  
FILING DATE: 08-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: YUN, Allen C.  
REGISTRATION NUMBER: 37,971  
REFERENCE/DOCKET NUMBER: TSUSAKI=1B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2889 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-09-024-429-4

Query Match 2.5%; Score 73.6; DB 3; Length 2889;

Best Local Similarity 47.0%; Pred. No. 3.9e-10;  
Matches 448; Conservative 0; Mismatches 479; Indels 27; Gaps 6;  
QY 1307 CGGCGATTTCAGAGGGCTTGAAGATATAAATTCCTTTATTTCAAGAGCTTGGTTGACTTA 1366  
DB 81 CGGGGACTTTGAGGGCTTGAAGGAGCTTCCCTACCTGGAGGAGCTCGGGGTCAACAC 140  
QY 1367 TGTGACCTGTATGCGCTCTTTAATGCTTGAAGGCAAGCGGCGGTATGCGGT 1426  
DB 141 CCTCTGGCTCATGCTCTTTCAGTCCCTTCTTCTGAGTCCCTTCTGAGGAGCGGTATGAT 194  
QY 1427 CAGCAGCTACCGGATGTCAATCGGCACTTGGGCAATAGGCGACTTTCGGGGAAGTCA 1486  
DB 195 CTCGGACTACTACCAATCTCCCTCCACGGGACCTTGGAGACTTCCACCG-----T 248  
QY 1487 TGTGCGCTGACGAAGCGGCGATTTCCGCGCTGCTGCTGATTTTATCTTCAACACACCTC 1546  
DB 249 GGACGAGGCGCCACGGCGGGGATGAAGGTGATCATTTGAGCTCGTCTGAAACACACCTC 308  
QY 1547 CACGACACGAAATGGGCGCAACGCTGCGCGCGCGGACCGGCTTTTCGACAAATTTCTA 1606  
DB 309 CATTTGACCACTTGTGTTCCAG-----GAGGCGAGGAGCGCAATAGCCCATGCGGA 362  
QY 1607 CTATATTTTCCCGACCGCGGATGCGGACCAATAGCAGCGCACCTTGGGCGAATCTT 1666  
DB 363 CT---GGTACGTGTGAGCGACACCCCGGAGAGTACAAAGGGGTTCGGGTATCTTTCAA 419  
QY 1667 CCCGACCGACACCGGGGCTTCTCGCACTTGAAGAGCGGACGCTGGGTGTTGAAGCAC 1726  
DB 420 GGACTTTGAACCTTCAACTTGAACCTTTGACCCCGTGGCCAGGCTTACTACTGACCG 479  
QY 1727 CTTCAATTCCTTCCAATGGACTTGAATTAACGCAACCCGTTGGGTATTCGGCGCAATGGC 1786  
DB 480 CTTCTACTGGCACCGCCGACCTCAACTGGGACAGCCCGAGGTGGAGAGGCCATCCA 539  
QY 1787 GGGCGAATCTGTTCTTTCGCACTTGGGCTTGAATCTCTGCTATGATGCGGTTC 1846  
DB 540 CCAGGTGATGTTCTTTCGGCGGACCTTGGGGGTGGAGCGCTTCGCTTGAAGCCATCCC 599  
QY 1847 CTTTATTTGGAACAAATGGGCAACGCTCGAAACCTTCCGCGAGCGC---ACGCCCT 1903

## RESULT 13

US-08-537-002A-5

; Sequence 5, Application US/08537002A

; Patent No. 5773282

; GENERAL INFORMATION:

; APPLICANT: TSUSAKI, Keiji

; APPLICANT: KUBOTA, Michio

; APPLICANT: SUGIMOTO, Toshiyuki

; TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR

; TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE

; NUMBER OF SEQUENCES: 17

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BROWDY AND NEIMARK

; STREET: 419 Seventh Street, N.W., Suite 300

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20004

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/537,002A

; FILING DATE: 29-SEP-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 260984/1994

; FILING DATE: 01-OCT-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP No. 5773282 yet received

; FILING DATE: 08-SEP-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: BROWDY, Roger L.

; REGISTRATION NUMBER: 25,618

; REFERENCE/DOCKET NUMBER: TSUSAKI=1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-628-5197

; TELEFAX: 202-737-3528

; TELEX: 248633

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3600 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: DOUBLE

600 CTACCTTACGAGCGGAGGAGACCTCTCTCGGAGAACCTCCCGAGACCACTTTCGAGCGGT 659  
1904 CATCCGCGGTTCAATGCCCTTATGCTATGCGCGCCCGCGTGTCTTCAATCCGA 1963  
660 GAAGCGCTGAGGAAGGCTTGGAGGAGCGCTACGGCCCGCGGAGATCTCTTCGCGCA 719  
1964 AGCATCTCCACCCCGACCAAGTCGTCCTCAATACATCGGCA--GGACGAATGCCAAAT 2020  
720 GGCACCAATGTGCGCGGAGGAGACCTCTCTCTTCTGCGGAGCGGAGCGGGTCCACAT 779  
2021 CGGTTACAAACCCCTTGCATAATGGCAATTTGTGTGGAACACCTTTCGCCACGCGCGGAGTCAA 2080  
780 GGCCTACAACTTCCCTCTGATGCCCGGATCTTCAATGGCCCTAAGCGCGGAGGAGCGGG 839  
2081 CTGCTTCATCAGGCGCTGACCTACGCCCAACCTTCCCGAGCATACCGCTCGGTCAA 2140  
840 TCCATTGAACCATGCTCAAGGAGGCGGAGGATCCCGAAACCCCGCAGTGGGCGCT 899  
2141 CTAGTCCGAGCCACGACGACATCGGCTGGACGTTTTTCCGATGAAGACGCGGATATCT 2200  
900 CTTCTCTCGCAACCAAGCGAGCTCACCTTGGAGAGGTCAAGGAGGAGCGGAGTT 959  
2201 GGGCATAGCGGCTACGACACCGCGCAATTCCTCAACCGCTTCTTCGTCAACCG 2254  
960 CATGTACGAGGCTTACGCGCCCGACCCCAAGTTCCGATCAACCTGGGATCCG 1013

QY	2081	CTGTGCTCCATCATAGGCGCTCACTACCGCACAAACCTGCGCCGAGCATACCGCCCTGGGCTCAA	214
Db	1380	TCCCATTTGAACCATCTCTAAGAGGCGGAGGGGATCCCGGAAACCGCCCACTGGGGCCCT	1439
QY	2141	CTAAGTCCGACGCCAGCAGCATATCGCTGGAGCTTTTCCGATGAAGAGCGGCATATCT	2200
Db	1440	CTTCTCTCGCGAACCAACGACGAGGCTCACCTCGGAGAGAGTCAACGAGGAGGAGCGGAGTT	1499
QY	2201	GGGCATAAGCGGCTACGACACACCGCAATTCCTCAACCGCTCTTCCTCAACCG	2254
Db	1500	CATGTACGAGGCGCTACGCCCGCCGACCCCAAGTTCGCAATCAACCTGGGATCCG	1553

RESULT 14

US-08-863-010-5

; Sequence 5, Application US/08863010

; Patent No. 6087146

GENERAL INFORMATION:

APPLICANT: TSUSAKI, Keiji

APPLICANT: KUBOTA, Michio

APPLICANT: SUGIMOTO, Toshiyuki

TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR

TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

ADDRESSEE: BROWDY AND NETMARK

STREET: 419 Seventh Street, N.W., Suite 300

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/863,010

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/537,002

FILING DATE: 29-SEP-1995

APPLICATION NUMBER: JP 260984/1994

FILING DATE: 01-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP No. 6087146 yet received

FILING DATE: 08-SEP-1995

ATTORNEY/AGENT INFORMATION:

NAME: BROWDY, Roger L.

REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER: TSUSAKI=1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-628-5197

TELEFAX: 202-737-3528

TELEX: 248633

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 3600 base pairs

TYPE: nucleic acid

STRANDEDNESS: DOUBLE

TOPOLOGY: linear

MOLECULE TYPE: genomic DNA

ORIGINAL SOURCE:

ORGANISM: *Thermus aquaticus*

INDIVIDUAL ISOLATE: ATCC 33923

FEATURE:

NAME/KEY: 5'UTR

LOCATION: 1..540

IDENTIFICATION METHOD: E

NAME/KEY: mat peptide

LOCATION: 541..3429

IDENTIFICATION METHOD: S

NAME/KEY: 3'UTR  
LOCATION: 3430..3600  
IDENTIFICATION METHOD: E  
US-08-843-010-5

Query Match 2.5%; Score 73.6; DB 3; Length 3600;  
Best Local Similarity 47.0%; Pred. No. 4.3e-10;  
Matches 448; Conservative 0; Mismatches 479; Indels 27; Gaps 6;

QY	1307	CGGGGATTGAGGCTTGAAGATAAAATTCCTATTTCCTCAAGAGCTTGTTGACTTA	1366
DB	621	CGGGGACTTTGAGGCTTGAGCGGAAGCTTCCTTACTCTGAGGAGCTCGGGGTCAACAC	680
QY	1367	TCTGCACTGATGCGCTGTTTAAATGCTTGAAGCAAAAGCGAGCGCGCTATGCGGT	1426
DB	681	CTCTGGCTCATGCGCTTCCTCCAGTCCG-----CTTGAGGAGCAACGGGTACGATAT	734
QY	1427	CAGCAGCTACCGCATGTCAATCGGCATCTGGCACAATGCGGACTTGCCTGAGTCAT	1486
DB	735	CTCCGACTACTACCAAGATCTCCCGCTCCACGGGACCTGAGGAGCTTCCACG-----T	788
QY	1487	TGCTGCGCTGACCAAGCGCGCAITTCGCGCTGCTGCAITTTTATCTTCAACCAACATC	1546
DB	789	GGACGAGGCCACCGCGCGGGGATGAAGTGATCATTCAGCTCGTCTGGAACCAACATC	848
QY	1547	CAACGAACAGCAATGGGCGCAACGCTGCGCGCGCGGACCGCTTTTCGACAATTTCTA	1606
DB	849	CATTGACCACTTGGTTTCCAG-----GAGCGGAGGAAGCGCAATAGCCCATGCGGA	902
QY	1607	CTATATTTTCCCGACCGCGGATGCCCGACCAATACGACCGCACCTTCGCGCAATCTT	1666
DB	903	CT---GGTACGTGTGGAGGACACCCCGGAGAGTACAAAGGGTCCGGTCACTTCA	959
QY	1667	CCCCGACGACACCCGCGGGGCTTCGGAACCTGGAAGACGAGCGTGGGTGTGAGAC	1726
DB	960	GGACTTTGAAACCTCAACTGACCTTTGACCCCGTGCCCAAGGCTTACTACTGCAACG	1019
QY	1727	CTTCAATCTTCCAAATGGGACTTCAATACAGCAACCGCTGGGTATTCCGCGCAATGGC	1786
DB	1020	CTTCTACTGGACACGCGGACTCAATCTGGACAGCCCCCGAGGCTTGGAGAGGCAATCA	1079
QY	1787	GGGCAAAATGCTTCTTCCCAATCTGGCGCTTGACATCTGCTGATGCGGTGTC	1846
DB	1080	CCAGGTGATGTTCTTCTGGCGGAGCTGGGGGTGACGCTTCCGCTGGAGCGCATCC	1139
QY	1847	CTTTATTTGGAACAAATGGGAGCAAGCTGCAAAACCTGCGGAGGCGC---ACGCGCT	1903
DB	1140	CTACCTCTACGAGCGGAGGAGGACCTCTCGGAGAACCTCCCGAGACCATGAGCGGT	1199
QY	1904	CATCGCGGCTTCAATGCGTTATGCGTATTCGCGCGCGCGCGCTGTTCTTCAATCCGA	1963
DB	1200	GAGCGCTGAGGAAGGCTTGGAGAGCGCTACGCGCGCGGGAAGATCTCTCGCGCA	1259
QY	1964	AGCCTGCTGACCGGACCAAGTGTCCATACATCGGCA---GGACGATGCCAAT	2020
DB	1260	GGCCCAACATGTGCGCGGAGGACCTCCCTTCTGCGGAGCGGGAGCGGCTCCCAT	1319
QY	2021	CGGTTACACCCCTTGAATGCAATGTTGTGGAACACCTTTCGCGCGCGGAGTCA	2080
DB	1320	GACCTACACTTCCCTCTGATGCGCGGATCTTCATGGCCCTAAGCGCGGAGCGGG	1379
QY	2081	CTGCTTCCATAGGCGCTGACCTACCGCAACCTGCGCGGAGCATACCGCTCGGTCA	2140
DB	1380	TCCCATTTGAACCTGCTCAAGGAGCGGAGGAGTCCCGAAACCGCGGAGTGGCGCT	1439
QY	2141	CTACGTCGCGGACGACGACATCGGCTGGAGCTTTGCGGATGAAGACGCGGATATCT	2200
DB	1440	CTTCTTCCGAAACCAAGCAGCTCACCCTGAGAGGTACGAGGAGGAGCGGAGTT	1499
QY	2201	GGGCATAAGCGCTACGACACCGCAATTCCTCAACCGCTTCTTGGTCAACCG	2254
DB	1500	CATGTACGAGGCTTACGCGCGGAGGAGGAGTTCGCTCATCACTCGGGGATCCG	1553

RESULT 15  
US-09-024-429-5  
Sequence 5, Application US/09024429  
Patent No. 6165768  
GENERAL INFORMATION:  
APPLICANT: TSUSAKI, Keiji  
APPLICANT: SUGIMOTO, Toshiyuki  
TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR  
CONVERTING MALTOSE INTO TRISALOSE  
TITLE OF INVENTION: 17  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/024,429  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/537,002  
FILING DATE: 29-SEP-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 260984/1994  
FILING DATE: 01-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 255829/1995  
FILING DATE: 08-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: YUN, Allen C.  
REGISTRATION NUMBER: 37,971  
REFERENCE/DOCKET NUMBER: TSUSAKI=1B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3600 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: DOUBLE  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Thermus aquaticus  
INDIVIDUAL ISOLATE: ATCC 33923  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 1..540  
IDENTIFICATION METHOD: E  
NAME/KEY: mat peptide  
LOCATION: 541..3429  
IDENTIFICATION METHOD: S  
NAME/KEY: 3'UTR  
LOCATION: 3430..3600  
IDENTIFICATION METHOD: E  
US-09-024-429-5

Query Match 2.5%; Score 73.6; DB 3; Length 3600;  
Best Local Similarity 47.0%; Pred. No. 4.3e-10;  
Matches 448; Conservative 0; Mismatches 479; Indels 27; Gaps 6;  
QY 1307 CGGGGATTGAGGCTTGAAGATAAAATTCCTATTTCCTCAAGAGCTTGTTGACTTA 1366  
DB 621 CGGGGACTTTGAGGCTTGAGCGGAAGCTTCCTTACTCTGAGGAGCTCGGGGTCAACAC 680

QY 1367 TGTGACCTGATGCGCTGTTAAATGCGCTGAGGCAAAAGCGAGCGGCTATGCGGT 1426  
 Db |||||  
 QY 681 CCTTGGCTCATGCTCTTCCAGTCCC-----CTTGAGGAGACGGGTACGATAT 734  
 Db |||||  
 QY 1427 CAGCAGCTACCGGAGTCAATTCGGCACTGGGCAATAGCGGACTTGGCGGAAGTCA 1486  
 Db |||||  
 QY 735 CTCGCACTACTACAGATCTCCCGCTCCACGGGACCTGGAGGACTTCAACG----- 788  
 Db |||||  
 QY 1487 TGTGCGCTGACAGCGGCAATTCGCGCTGCTGATTTTATCTTCAACCAACACCTC 1546  
 Db |||||  
 QY 789 GGAGGAGGCGGCGGCGGAGTGAAGTGAATCTGAGCTCGTCTGACCAACACCTC 848  
 Db |||||  
 QY 1547 CAACGAACAGAAATGGGCGCAACGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGG 1606  
 Db |||||  
 QY 849 CATTGACCACTTGGTTCCAG-----GAGGCGAGGAAGCGCAATAGCCCATCGGGA 902  
 Db |||||  
 QY 1607 CTATATTTTCCCGGCGGCGGAGTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1666  
 Db |||||  
 QY 903 CT---GGTACGTGTGAGCGACACCGCGGAGAGTACAAGGGGCTCCGGGTCTTTCAA 959  
 Db |||||  
 QY 1667 CCGGACCAAGCAGCGGCGGCTTCTCGCAACTGGAAGACGACGCTGGGTGTGGAACGAC 1726  
 Db |||||  
 QY 960 GGAATTTGAACCTTCAACTGGACCTTTGACCCCGTGGCGCAAGGCGCTACTACTGGCA 1019  
 Db |||||  
 QY 1727 CTTCAATTCCTTCAATGGGCACTTGAATTAAGCAACCGTGGGTATTCGGGCGAATGCG 1786  
 Db |||||  
 QY 1020 CTTCTACTGCGACCACTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1079  
 Db |||||  
 QY 1787 GGGCGAAATGCTGCTTCTTCCAACTTGGGCGTGTGACATCTCGGCTGATGCGGTTGC 1846  
 Db |||||  
 QY 1080 CAGGTCAATGTTCTTCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1139  
 Db |||||  
 QY 1847 CTTTATTGGAAACAAATGGGCAAGCTGGGCAAACTTGGCGGCGGCGGCGGCGGCGGCGG 1903  
 Db |||||  
 QY 1140 CTACCTCTACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1199  
 Db |||||  
 QY 1904 CATCGCGGCTTCAATGCGGCTTATGGTATTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1963  
 Db |||||  
 QY 1200 GAAGCGCTGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1259  
 Db |||||  
 QY 1964 AGCATGCTCCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2020  
 Db |||||  
 QY 1260 GGCACCACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1319  
 Db |||||  
 QY 2021 CGGTTACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2080  
 Db |||||  
 QY 1320 GGCCTACAACTTCCCGCTGATGCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1379  
 Db |||||  
 QY 2081 CTTGCTCCATCAGCGGCTGACCTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2140  
 Db |||||  
 QY 1380 TCCCATTTGAACCTATGCTCAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1439  
 Db |||||  
 QY 2141 CTAAGTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2200  
 Db |||||  
 QY 1440 CTTCTCCGCAACCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1499  
 Db |||||  
 QY 2201 GGGCATAGGCGGCTAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2254  
 Db |||||  
 QY 1500 CATGTAGGCGGCTAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1553  
 Db |||||

Search completed: November 8, 2003, 17:35:05  
 Job time : 214 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 16:01:07 ; Search time 886 Seconds  
(without alignments)

10494.866 Million cell updates/sec

Title: US-09-843-007A-1

Perfect score: 2914  
Sequence: 1 GAGTTTCGTTCCGACAC.....CGTTTCAGCGCATTTGCG 2914

Scoring table: IDENTITY NUC

Gapop 10\_0, Gapext 1.0

Searched: 2141354 seqs, 1595478879 residues

Total number of hits satisfying chosen parameters: 4282708

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:

1: /cgn2\_6/ptodata/2/pubna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubna/PTC\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubna/PTCUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubna/US09C\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubna/US10\_NEW\_PUB.seq.\*  
16: /cgn2\_6/ptodata/2/pubna/US60\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/2/pubna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2914	100.0	2914	9	US-09-843-007-1
2	169	5.8	759	9	US-09-815-242-6076
3	161.4	5.5	882	9	US-09-815-242-9839
4	142.6	4.9	801	9	US-09-815-242-6759
5	142.6	4.9	6415	10	US-09-070-927A-275
6	120.4	4.1	3303	9	US-09-815-242-7764
7	119	4.1	1716	14	US-10-156-761-2790
8	119	4.1	9025608	14	US-10-156-761-1
9	118	4.0	483	10	US-09-974-300-399
10	111.8	3.8	1794	10	US-09-738-626-2527
11	111.8	3.8	3303400	10	US-09-738-626-1
12	111	3.8	1749	14	US-10-156-761-7374
13	70.8	2.4	1599	14	US-10-156-761-1713
14	70.8	2.4	9025608	14	US-10-156-761-1
15	65	2.2	1589	10	US-09-974-300-657
16	59.8	2.1	1782	14	US-10-061-269-13

17	56.8	1.9	1704	14	US-10-156-761-6284
18	52	1.8	1683	10	US-09-974-300-655
19	47.2	1.6	520	14	US-10-184-644-332
20	47.2	1.6	520	14	US-10-184-644-332
21	47.2	1.6	1848	14	US-10-156-761-5957
22	46.8	1.6	1704	14	US-10-061-269-15
23	45.2	1.6	1703	10	US-09-974-300-644
24	44.2	1.5	1659	12	US-10-223-277-7
25	44.2	1.5	1659	12	US-10-223-277-8
26	43.4	1.5	6142	13	US-10-014-436-1
27	42.8	1.5	1344	14	US-10-156-761-6042
28	42.6	1.5	1653	9	US-09-815-242-9889
29	42.6	1.5	30246	8	US-08-781-986A-56
30	42.4	1.5	1521	14	US-10-156-761-6290
31	42.4	1.5	77536	12	US-09-940-316B-1
32	41.8	1.4	5418	14	US-10-156-761-5959
33	41.6	1.4	933	14	US-10-156-761-2119
34	41.6	1.4	933	14	US-10-156-761-2300
35	41.6	1.4	963	14	US-10-156-761-686
36	41.4	1.4	1872	12	US-10-084-843-17
37	41.4	1.4	1872	12	US-10-193-002-17
38	41.4	1.4	1872	12	US-10-098-732A-1
39	41.4	1.4	2151	14	US-10-127-032-72
40	41.4	1.4	4935	14	US-10-156-761-5059
41	41.2	1.4	2586	10	US-09-738-626-456
42	40.6	1.4	1785	10	US-09-738-626-2326
43	40.6	1.4	2325	14	US-10-156-761-2691
44	40.6	1.4	3033	14	US-10-212-219-5
45	40.2	1.4	1389	14	US-10-156-761-5431

#### ALIGNMENTS

#### RESULT 1

US-09-843-007-1

; Sequence 1, Application US/09843007

; Patent No. US20020092040A1

; GENERAL INFORMATION:

; APPLICANT: Rosemann, Jens

; Butcher, Volker

; Weish, Thomas

; TITLE OF INVENTION: DNA SEQUENCES ENCODING ENZYMES CAPABLE

; OF FACILITATING THE SYNTHESIS OF LINEAR

; ALPHA-1,4 GLUCANS IN PLANTS, FUNGI AND

; MICROORGANISMS

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: James F. Haley, Jr., c/o Fish & Neave

; STREET: 1251 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10020

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC Compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/843,007

; FILING DATE: 26-Apr-2001

; CLASSIFICATION: <Unknown>

; ERROR APPLICATION DATA:

; APPLICATION NUMBER: 08/737,752

; FILING DATE: <Unknown>

; APPLICATION NUMBER: DE P 44 47 388.5

; FILING DATE: 22-DEC-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: James F. Haley, Jr., Esq.

; REGISTRATION NUMBER: 27,794

; REFERENCE/DOCKET NUMBER: GFB-1

; TELECOMMUNICATION INFORMATION:

Sequence 6284, Ap  
Sequence 655, App  
Sequence 332, App  
Sequence 332, App  
Sequence 5957, Ap  
Sequence 15, Appl  
Sequence 644, App  
Sequence 7, Appl  
Sequence 8, Appl  
Sequence 1, Appl  
Sequence 6042, Ap  
Sequence 9889, Ap  
Sequence 56, Appl  
Sequence 6290, Ap  
Sequence 1, Appl  
Sequence 5959, Ap  
Sequence 2119, Ap  
Sequence 2300, Ap  
Sequence 686, App  
Sequence 17, Appl  
Sequence 17, Appl  
Sequence 1, Appl  
Sequence 72, Appl  
Sequence 5059, Ap  
Sequence 456, App  
Sequence 2326, Ap  
Sequence 2691, Ap  
Sequence 5, Appl  
Sequence 5431, Ap

TELEPHONE: (212) 596-9000  
TELEFAX: (212) 596-9090

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2914 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Neisseria polysacchara

IMMEDIATE SOURCE:

LIBRARY: genomic library in pBluescriptII SK

CLONE: pNB2

FEATURE:

NAME/KEY: CDS

LOCATION: 957..2867

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-843-007-1

Query Match 100.0%; Score 2914; DB 9; Length 2914;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2914; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GAGTTTGGCTTCCCGAACCGAAGCGTGTGCTGAGCCGACACCTGTCCGCGCAAGGCGG 60  
1 GAGTTTGGCTTCCCGAACCGAAGCGTGTGCTGAGCCGACACCTGTCCGCGCAAGGCGG 60

61 CTGACCGCCCTTTTCCCGAACCGAAGCGTGTGCTGAGCCGACACCTGTCCGCGCAAGGCGG 120  
61 CTGACCGCCCTTTTCCCGAACCGAAGCGTGTGCTGAGCCGACACCTGTCCGCGCAAGGCGG 120

121 GCTTTGAGCGTGGCAGAAACGAAAGTCAAGCGTCTTCCGCGCTTTGCGGCAATCAGGAT 180  
121 GCTTTGAGCGTGGCAGAAACGAAAGTCAAGCGTCTTCCGCGCTTTGCGGCAATCAGGAT 180

181 APTTTCAGATGTCGCGCGCGAGTCTCTCATCTGTTTCAGAGCGGATACGATTTCTTCT 240  
181 APTTTCAGATGTCGCGCGCGAGTCTCTCATCTGTTTCAGAGCGGATACGATTTCTTCT 240

241 TCGCGCGCGTGGCGTGAACCTCATGATTCAGAGCGGCGGCGGATGCGGTTTTTGA 300  
241 TCGCGCGCGTGGCGTGAACCTCATGATTCAGAGCGGCGGCGGATGCGGTTTTTGA 300

301 GCATGCGCGCGCGCGCGCGGCGGTTTCGCGGAAAGTCTCGATATCGATAATG 360  
301 GCATGCGCGCGCGCGCGCGGCGGTTTCGCGGAAAGTCTCGATATCGATAATG 360

361 TCGCGCGCGCGGCTTTCAATCAGCGAGTCCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 420  
361 TCGCGCGCGCGGCTTTCAATCAGCGAGTCCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 420

421 GCGAAGCGCGCGCTTCCGATCGCTGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 480  
421 GCGAAGCGCGCGCTTCCGATCGCTGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 480

481 GCGTTCGCGGAGCGGCTTCCGATCGCTGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 540  
481 GCGTTCGCGGAGCGGCTTCCGATCGCTGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 540

541 AAATCGCGCGGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 600  
541 AAATCGCGCGGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 600

601 GAAAGTACCGCGGATCGCGCGGCGGCAAGAGCGGATTTTGGTGGTTCGCTTCGCG 660  
601 GAAAGTACCGCGGATCGCGCGGCGGCAAGAGCGGATTTTGGTGGTTCGCTTCGCG 660

661 ATAAAGCGTGTGAGCGGTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 720  
661 ATAAAGCGTGTGAGCGGTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGT 720

QY 721 GGTAAAGACCAATTATAGCAAGGACAGCGCATGTTTCAGACGCGATTTCTGTGGCG 780  
DB 721 GGTAAAGACCAATTATAGCAAGGACAGCGCATGTTTCAGACGCGATTTCTGTGGCG 780  
QY 781 CGGCTTGATATGAATCAAGCAGCATCCGCTATATCGAATCGAGACTTTGGCAAGCCCTG 840  
DB 781 CGGCTTGATATGAATCAAGCAGCATCCGCTATATCGAATCGAGACTTTGGCAAGCCCTG 840  
QY 841 TCTTTTCTAGTCAGTCGCGAGTCTTTCGAGTATGATTCAGCAGCAGCCCTTACAGCGAT 900  
DB 841 TCTTTTCTAGTCAGTCGCGAGTCTTTCGAGTATGATTCAGCAGCAGCCCTTACAGCGAT 900  
QY 901 TTGAGGATACGCGCAGACCGCGCTCGAATTCGAAATCGAGCAGCGATCATGT 960  
DB 901 TTGAGGATACGCGCAGACCGCGCTCGAATTCGAAATCGAGCAGCGATCATGT 960  
QY 961 TGACCCCGACGCAAGTCTGTTGATTTTACAGTACTCTCAAAACAGCATCTTGACA 1020  
DB 961 TGACCCCGACGCAAGTCTGTTGATTTTACAGTACTCTCAAAACAGCATCTTGACA 1020  
QY 1021 TCTACAGCCCGAAACAGCGCGCGCATCGAATAATCGAAGACTTGGCGGCGATTTTCCG 1080  
DB 1021 TCTACAGCCCGAAACAGCGCGCGCATCGAATAATCGAAGACTTGGCGGCGATTTTCCG 1080  
QY 1081 GCGCATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
DB 1081 GCGCATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
QY 1141 ACGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200  
DB 1141 ACGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200  
QY 1201 AACGCAATCATCTCTTAAAGATATCGATATCGATATCGATATCGATATCGATATCGATAT 1260  
DB 1201 AACGCAATCATCTCTTAAAGATATCGATATCGATATCGATATCGATATCGATATCGATAT 1260  
QY 1261 TGTCCAAACAAAGTCCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320  
DB 1261 TGTCCAAACAAAGTCCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320  
QY 1321 GCTTGAAGAGATAAATTCCTTTTCAAGAGCTTGGTTGACTTATCTGACCTGATGCT 1380  
DB 1321 GCTTGAAGAGATAAATTCCTTTTCAAGAGCTTGGTTGACTTATCTGACCTGATGCT 1380  
QY 1381 CGCTGTTTAAATGCTTGAAGGCAAAAGCGAGCGGCTATGCGTACGAGCTACCGG 1440  
DB 1381 CGCTGTTTAAATGCTTGAAGGCAAAAGCGAGCGGCTATGCGTACGAGCTACCGG 1440  
QY 1441 ATGTCATATCCGCGACTGCGGCAATAGCGGACTTGGCGGAGTCAATGCTGCTGCTGCTGCT 1500  
DB 1441 ATGTCATATCCGCGACTGCGGCAATAGCGGACTTGGCGGAGTCAATGCTGCTGCTGCTGCT 1500  
QY 1501 AAGCGCGATTTCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560  
DB 1501 AAGCGCGATTTCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560  
QY 1561 GGGCGCAAGCTGCGCGCGCGCGCGCTTTCGACAAATTTCTACTATATTTTCCCG 1620  
DB 1561 GGGCGCAAGCTGCGCGCGCGCGCGCTTTCGACAAATTTCTACTATATTTTCCCG 1620  
QY 1621 ACCGCGGATGCGCGCGCGCGCGCGCTTTCGACAAATTTCTACTATATTTTCCCG 1680  
DB 1621 ACCGCGGATGCGCGCGCGCGCGCGCTTTCGACAAATTTCTACTATATTTTCCCG 1680  
QY 1681 GGGCGCGCTTTCGCGCACTGGAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740  
DB 1681 GGGCGCGCTTTCGCGCACTGGAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740  
QY 1741 AATGGAATTTGAAATACAGCAACCGCGGCTTTCGCGCAATTTCCGCGCAATTTCCGCGCAAT 1800  
DB 1741 AATGGAATTTGAAATACAGCAACCGCGGCTTTCGCGCAATTTCCGCGCAATTTCCGCGCAAT 1800

QY 1801 TCCTTGCCAACTTGGGGGTTGACATCCCTGCGTATGATGCGGTTGCTTATTTTGAAC 1860  
DB 1801 TCCTTGCCAACTTGGGGGTTGACATCCCTGCGTATGATGCGGTTGCTTATTTTGAAC 1860  
QY 1861 AAATGGGACAGCTTGGGAAACCTGCGGAGGCGAGCCCTCATCGCGCGTTCAATG 1920  
DB 1861 AAATGGGACAGCTTGGGAAACCTGCGGAGGCGAGCCCTCATCGCGCGTTCAATG 1920  
QY 1921 CCGTATGCGTATGCGCGCGCGCGGTTGTTCTTCAAATCGAAGCCATGCTCAACCCG 1980  
DB 1921 CCGTATGCGTATGCGCGCGCGCGGTTGTTCTTCAAATCGAAGCCATGCTCAACCCG 1980  
QY 1981 ACCAAGTGGTCCATACATCGGAGGAGGAGATGCAATCGGTTTACAAACCCCTGCAA 2040  
DB 1981 ACCAAGTGGTCCATACATCGGAGGAGGAGATGCAATCGGTTTACAAACCCCTGCAA 2040  
QY 2041 TGGCATTTCTTGGGAACACCCCTTGCACGCGGAGTCAACCTGCTCCATCAGGCGCTGA 2100  
DB 2041 TGGCATTTCTTGGGAACACCCCTTGCACGCGGAGTCAACCTGCTCCATCAGGCGCTGA 2100  
QY 2101 CCTACCGCACACCTGCGGAGGAGTACGCGTGGTCAACTACCTCGCGAGCCAGGAG 2160  
DB 2101 CCTACCGCACACCTGCGGAGGAGTACGCGTGGTCAACTACCTCGCGAGCCAGGAG 2160  
QY 2161 ACATCGGCTGAGGTTTCCGATGAAGAGCGGCGATATCTGGGCATAGCGGCTACGAC 2220  
DB 2161 ACATCGGCTGAGGTTTCCGATGAAGAGCGGCGATATCTGGGCATAGCGGCTACGAC 2220  
QY 2221 ACCGCAATTCCTCAACCGCTTCTTGGTCAACCGTTTTCGAGGAGGCTTGGTGGCG 2280  
DB 2221 ACCGCAATTCCTCAACCGCTTCTTGGTCAACCGTTTTCGAGGAGGCTTGGTGGCG 2280  
QY 2281 TACCGTTCCATACAAACCCAGCAGGAGTCCGCTGCTAGTGGTACAGCGCGGCTAT 2340  
DB 2281 TACCGTTCCATACAAACCCAGCAGGAGTCCGCTGCTAGTGGTACAGCGCGGCTAT 2340  
QY 2341 TGGTGGCTGCGGCAAGAGATCCCAACCGCTTGAACCGGATCAAACTTTGTACAGCA 2400  
DB 2341 TGGTGGCTGCGGCAAGAGATCCCAACCGCTTGAACCGGATCAAACTTTGTACAGCA 2400  
QY 2401 TTGCTTTGATGACGCGGCTGCGCTGATTTACTAGGCGAGGAGTGGTACGCTCA 2460  
DB 2401 TTGCTTTGATGACGCGGCTGCGCTGATTTACTAGGCGAGGAGTGGTACGCTCA 2460  
QY 2461 ATGAGGACGACTGCTGCAAGACAGCAATAAGAGCGAGCAGACCGCTTGGGCGCACCGTC 2520  
DB 2461 ATGAGGACGACTGCTGCAAGACAGCAATAAGAGCGAGCAGACCGCTTGGGCGCACCGTC 2520  
QY 2521 CCGGCTAACAGAGCCCTGTAGCGGACAGCAAGATCCGTCGACCGCGCGGCGCAA 2580  
DB 2521 CCGGCTAACAGAGCCCTGTAGCGGACAGCAAGATCCGTCGACCGCGCGGCGCAA 2580  
QY 2581 TCTATCAGGCTTGGCCATATGATTGCGCTCGCCAAAGCAATCGCGCTTTCGACGGCG 2640  
DB 2581 TCTATCAGGCTTGGCCATATGATTGCGCTCGCCAAAGCAATCGCGCTTTCGACGGCG 2640  
QY 2641 GCAGGCTGTTATCAATCAACAAACAGCAATCATCTGGCTACATCCGCAATG 2700  
DB 2641 GCAGGCTGTTATCAATCAACAAACAGCAATCATCTGGCTACATCCGCAATG 2700  
QY 2701 CCGTTTTCGATTCGCTTACGAGGATATCCGCAACCGTTTACCGGCGCATACCTGCG 2760  
DB 2701 CCGTTTTCGATTCGCTTACGAGGATATCCGCAACCGTTTACCGGCGCATACCTGCG 2760  
QY 2761 AAGCCATGCGCTTCAAGGCGCAGCCTCATCGTGGCGAAAATGTCAGCCTGATTCAG 2820  
DB 2761 AAGCCATGCGCTTCAAGGCGCAGCCTCATCGTGGCGAAAATGTCAGCCTGATTCAG 2820  
QY 2821 ATTGAGGCTTACGCGCTATCAGGCTCATGTTGGTTCGAATTCGCTGACGAGCTTCCCA 2880  
DB 2821 ATTGAGGCTTACGCGCTATCAGGCTCATGTTGGTTCGAATTCGCTGACGAGCTTCCCA 2880  
QY 2881 AATGCGCTGTAACCGCTTTCAGACGCGCAATTTGCG 2914

DB 2881 AATGCGCTGTAACCGCTTTCAGACGCGCAATTTGCG 2914

RESULT 2

US-09-815-242-6076/c  
; Sequence 6076, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/131,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14310  
; SOFTWARE: fastseq for Windows Version 4.0  
; SEQ ID NO 6076  
; LENGTH: 759  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(759)  
US-09-815-242-6076

Query Match 5.8%; Score 169; DB 9; Length 759;  
Best Local Similarity 53.8%; Pred. No. 1.1e-42;  
Matches 371; Conservative 0; Mismatches 315; Indels 3; Gaps 1;

QY 1 GAGTTTTCGTTCCGACCGAAGCTGATGCTTGGAGCGGACACCTGTCCGCGAAGCGG 60  
DB 695 GACGCTTTTTCACCGCACCAAAAGTTGCGCGGAGCCAAATATCTTACAGCCGAGCA 636  
QY 61 CTGACCGCGCCCTTTTGCCCATCGCATCGTAACTCGGTTTGGTGGCAA---GCTCT 117  
DB 635 GAATATACCGCAGTTTTCGATCGCATCGTATATTTGACGATCGGCATATCTGCTCC 576  
QY 118 TCGCTTTGAGCGTGGCGAAGCAAGTACAGCGCTTCCGCGCTTTCGCGCATCACC 177  
DB 575 TGCATCTCAGGCTGCGCGCAAGCAACGTCAGCATCGCTGGTACTCTTTGGGCGATCAGC 516  
QY 178 GCAATTTTGCAGATGTCGCGCGCGAGTCTTCCATCTGTTTCAGACGCGCATACGATTCT 237  
DB 515 GCAATCTTAGAATATCGCGTTCGAAGATTGCAATTTTTCGACAGCGGCAATGATTCT 456  
QY 238 TCTTTCGCGCGGCTGCGGTGAACCTCATGATTCAGACAGCGGCGCGCATCGGTTTCT 297  
DB 455 TCGGCTTCGCGGCTTTTATGGAAGTCAATGTTTGGACATGACTTTCACATCATGCGCG 396  
QY 298 TGAGCATCGCCACGCGCGCGGAGCGGCTTTCGCGGAGAAAGCTCGATATCGATA 357

Db 395 TGGCGGTAGGCGAAGCGTTCTTTTAACTGATCATCAGCGGTAATAACTCCAGATCGATC 536  
Qy 358 ATGTCCGGCAGGCGCTTTTCAATCAGCGAGTCGAGCAGTTTCAAAATAATAATATCTCCGAA 417  
Db 335 ATATCAACAGGCGCGCTGTGATGCTGCACGATTGAGTGCATATAAGGCTCGGTGGAA 276  
Qy 418 CACGGGAACGAGCGCCCTTCCGCTCTGACGCTTAACAGCAGCGGCTTGTCCGGC 477  
Db 275 ATGCGCTGTCCCGCCCTTCTTTTGCACTGCGGAGGTAAACAGCAGCGTTTCTTGGC 216  
Qy 478 AGCGGTGCGGGAAGCGCTTGTGCTGTGCGGCAATATCTCGCGATGCTGCCCGCGCAATTC 537  
Db 215 ATGCTCTCAACGAGAAATTTTCTGCCCGCATGACAGACTCCCATTTGGAGAGGTCCGCA 156  
Qy 538 AAAAATCGCGCGAATCTCGACGATATCGAGGAGGCGAGTTTGTATTTGTCAAGTACG 597  
Db 155 TAGTGGTCAACAGCCATTCAGAAATCAAGTCCGCTTCAAGTACGAGAGCTTCG 96  
Qy 598 GCGGAAGTACGCGGCAATCGCGGCAAGCGGCAAGCGGCAAGCGGCAAGCGGCAAGCGGCT 657  
Db 95 GATTTCAGCGTGGCGATATCTTTCGCAATCAGCGAGAGCATGATTTTAGTGGCGCGGTA 36  
Qy 658 CGATAACGCTTTTTCAGCGTCAAGCT 686  
Db 35 CCAATGACGAGATCTTTTACAGTTACGCT 7

## RESULT 3

US-09-815-242-9839/c

; Sequence 9839, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-27

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 9839

; LENGTH: 882

; TYPE: DNA

; ORGANISM: Salmonella typhi

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)...(882)

US-09-815-242-9839

Query Match 5.5%; Score 161.4; DB 9; Length 882;  
Best Local Similarity 53.3%; Pred. No. 3.2e-40;  
Matches 364; Conservative 0; Mismatches 316; Indels 3; Gaps 1;

Qy 7 TCGCTTCCGGAACGAAAGCTGATGCTTTCAGCGCAACCTGTTCGCGCAAGCGGCTGACC 66  
Db 812 TGCTTTCACGCGCCAAAGCTGCGCGCAGAGCAAACTTCCCTTCGCGCAAGCGGCTGAAATG 753  
Qy 67 GCGCCCTTTTCCGCGCAGATCGTAACATCGTTTGTGTCGSC---AAGCTCTTTCGCT 123  
Db 752 ACACCTCTTTCGCGCATTTGACATAGTAATTACCGGAAGCGTGGGCTAATGTTCTGCTATC 693  
Qy 124 TTGAGCGTGGCAGAAAGCAAGTTCAGCACGCTTTCGCGGCTTTTTCGCGCATCAGCGCAATT 183  
Db 692 TCCAGCGTGGCAGTGAAGTAACTTAATACATCATGCTTTCGCGCATCAGCGCAATT 633  
Qy 184 TTGAGATGTCGCGCGCAGTCTCCATCTGTTTTCAGAGGCAATAGATTTCTTCTTTCG 243  
Db 632 TTGGAATATCCGCGCGAGTCTTTCGATTTTACGTAGCCGCGCAAGCAATTTCTCTGCG 573  
Qy 244 GCGCGCTGCGGTGAAGTCTCATGATTCAGAGGCGCGGCGGATGCGGCTTTTTCAGCA 303  
Db 572 GACGCGCTGCGTGAAGTCTGATTAAGTAACTTAAGTAACTTAAGTAACTTAAGTAACT 513  
Qy 304 TCGCGCAGCGCGCGCGGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCT 363  
Db 512 TATCGACGCTGCGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCT 453  
Qy 364 GCGAGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCG 423  
Db 452 ACCAGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCG 393  
Qy 424 AACGAGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCG 483  
Db 392 TGCTGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCG 333  
Qy 484 TCGCGCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 543  
Db 332 TCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGG 273  
Qy 544 TCGCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCG 603  
Db 272 TCCAGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCG 213  
Qy 604 AGTACGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCG 663  
Db 212 AGCTATTTGATGCTCTTCCCATCAACGACACGATATTTTGGGCTATCCCTTCGCAATG 153  
Qy 664 ACGGTGTTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCG 686  
Db 152 ATAAGATTTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCG 130

## RESULT 4

US-09-815-242-6759/c

; Sequence 6759, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 6759  
LENGTH: 801  
TYPE: DNA  
ORGANISM: Enterococcus faecalis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(801)  
US-09-815-242-6759

Query Match  
Best Local Similarity 4.9%; Score 142.6; DB 9; Length 801;  
Matches 353; Conservative 0; Mismatches 329; Indels 3; Gaps 1;

1 GAGTTTGGGTTCCCGAACGACGATGATCTTGAGCCGACACCTGTCGCGCAAGGGG 60  
734 GACGCTTGCTGGGCTGAGCCAAAGTAGTGGACCCGACACAGTTGACCTGTAGCGCA 675  
61 CTGACCGCCCTTTTCCCGACGACATCGTAACATCGTTGTTGGTGCAGCTCTTT- 119  
674 CTAATCATCCCTAACTGTCCCATTTGATGATGATGATGATGATGATGATGATG 615  
120 --CGCTTTGAGCGTGGCGAAGCAAGTACGACGCTCTTCCGCGCTTTGGCGCATACC 177  
614 TACATTTTCTGTTGCTGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 555  
178 GCATTTTTCAGATGTCGCGCGGAGTCTGCTGATCTGTTTTCAGCGCATAGATTTCT 237  
554 GCTATTTTCAAAATATCCGCTGACGCTATTCGATTTGGGTAACCGCGCAATCTCC 495  
238 TCTTGGCGCGGCTGGGTTGAACTCATGATTCAGAGCGGCGCGGATGCGGTTTCTT 297  
494 TCTTGGAGCGGTGTTTCTGAAAGTCTGATTTACAGAGCAATTTTAAATCCCATCTT 435  
298 TGAGCATGCGCCAGCGGCGCGGAGCGGCTGCTGATTTCCCGGAAAGTAAAGTAAAG 357  
434 TTAGCTTCGTGATTAACGCTATCTGCGCGGCGGATTCGCAATTAATCAATATCTAT 375  
358 ATGTCGGCGAGCGGCTTTCAATCAGCGAGTTCGAGGATTCAGGATTAATTAATCGT 417  
374 AATCCAGGCACTTTTTCATCTAATCATGATACAGCAAAATATTTTCTTCAGAA 315  
418 CACGGGAACGAGCGGCTTTCGCGATCGGCTGACGCTGACGCTGACGCTGACGCTG 477  
314 AAGCCATTTTCGCGGCTTCTTTTGTGACGAAAGTAAAGTAAAGTAAAGTAAAGTAA 255  
478 AGCGCTTCGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 537  
254 AACCGTTCCTAATTTGTTGTAAGATTAATACATACATACATACATACATACATACAT 195  
538 AAAAAATCGCGCGGAACTCGAGATATCGAGGCGAGGCTTTTGTGATTTGTTCAAGTAC 597  
194 TAATAGTCTAAACGCCATCCACCAATCAATCAATCAATCAATCAATCAATCAATCA 135  
598 GCGGAAGTACGCGGATCGCGGCGGAGCAAGCGGCGGAGTGTGCTGCTGCTGCTGCTG 657  
134 GCCAAATATCTCTCGGCTGTTGGAGCAAAATGGGCAACCAATTTTGGGATTTCCCTTCA 75  
658 CCGATACCGGTGTTTTCAGCGTCA 682  
74 CCTATTGCGACATTTTGTACTATAA 50

RESULT 5  
US-09-070-927A-275  
Sequence 275, Application US/09070927A  
Patent No. US20020120116A1  
GENERAL INFORMATION:  
APPLICANT: Charles A. Kunsch  
Steven Barash  
Patrick J. Dillon  
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
NUMBER OF SEQUENCES: 982  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/070,927A  
FILING DATE: 04-May-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/046,655  
FILING DATE: 1997-05-16  
APPLICATION NUMBER: 60/044,031  
FILING DATE: 1997-05-06  
APPLICATION NUMBER: 60/066,009  
FILING DATE: 1997-11-14  
ATTORNEY/AGENT INFORMATION:  
NAME: Kenley K. Hoover  
REGISTRATION NUMBER: 40,302  
REFERENCE/DOCKET NUMBER: PB369  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 275:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6415 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 275:  
US-09-070-927A-275

Query Match 4.9%; Score 142.6; DB 10; Length 6415;  
Best Local Similarity 51.5%; Pred. No. 9,9e-34;  
Matches 353; Conservative 0; Mismatches 329; Indels 3; Gaps 1;

1 GAGTTTGGGTTCCCGAACGACGATGATCTTGAGCCGACACCTGTCGCGCAAGGGG 60  
2727 GACGCTTGCTGGGCGGAGCCAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 2786  
61 CTGACCGCCCTTTTGGCCCATCGACATCGTAACATCGGTTTGGTGGCAAGCTCTTT- 119  
2787 CTAATCATCCCTAACTGCTCCCATTCGATGATGATGATGATGATGATGATGATG 2846  
120 --CGCTTTGAGCGTGGCGAAGCAAGTACGACGCTTTCGCGCTTTCGCGCATACC 177  
2847 TACATTTTCAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2906  
178 GCATTTTTCAGATGTCGCGCGCGGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 237  
2907 GCTATTTTCAAAATATCTGCTGAGCGATTTGCGATTTGCGGTAAACCGCGCAATCTCC 2966  
238 TCTTGGCGCGGCTGCGGTGAACTCATGATTTGCGAGCAGCGGCGGCGGATGCGGTTTT 297  
2967 TCTTGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3026

```

QY 298 TGAGCATGGCCACGGCGCCGCGAGCGGGTTTCCGCGGAAAAAGTCGATATCGATA 357
Db 3027 TTAGCTTCGTGAATTAACGTATCTGCCGCCAGCGGATTCGCAATAATTAATCAATCCAA 3086
QY 358 ATGTCGGGAGGCGGCTTCAATCAGCGAGTCGAGCAGTTCAAAATAAATAATCGTCCGAA 417
Db 3087 AAATCCACCGCACCTTTTACTAATTCAATGATACATGCAAAATAATTTCTTCGGA 3146
QY 418 CACGGGAACGAGCGGCTTCGCATCGCGTCTGAACGTAAACAGCAGCGGCTTGTGCGGC 477
Db 3147 AAAGCCATTTCCGCGCTTCTTTTGTGTACGAAAGTTAAACATAACGGTTTTTCCGCC 3206
QY 478 AGCGCGTCGGGAGCGTCTGGGTGTGGCGCAATACTTCGCGCATGCTGCCGCGCATTC 537
Db 3207 AACGTTTCCATTAATCTGTGTGAAGATTACATACATCGGAAATACAGGACATTTTCA 3266
QY 538 AAAAAATCGCGGCGAACTCGACGATATCGAAGCGCAGGTTTTTGTGATGTAAGTAAG 597
Db 3267 TAATAGTCTAAAGCCATTCCACCATAATCAATCTAGTGTGTGGCTGCGCGTTGCTTGG 3326
QY 598 GCGGAAGTAGCGGCGCATCGCGGCGACAGCGGCGAGCGGATTTTGTGCGGTCCGCTT 657
Db 3327 GCCAAATATATCTCGGTGTGTGGAGCAACATGGGCGACCAATTTTGGGATTTCCCTTCA 3386
QY 658 CCGATAACGGTGTGTTTTGACGGTCA 682
Db 3387 CCTATTCCACATTTTGGACTATAA 3411

```

RESULT 6

```

US-09-815-242-7764
; Sequence 7764, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7764
; LENGTH: 3303
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)....(3303)
US-09-815-242-7764

```

Query Match 4.1%; Score 120.4; DB 9; Length 3303;

```

Best Local Similarity 47.9%; Pred. No. 8e-27;
Matches 463; Conservative 0; Mismatches 476; Indels 27; Gaps 3;
QY 1307 CGCGAATTGAAGGCGCTTGAAGATATAAATTCCTTATTTTCAAGAGCTTGGTTTCACTTA 1366
Db 111 CGCGCACTTCGCGGCGTGTATCGAGAGCTCGACTACATCGCCGACCTCGCGGTGAACAC 170
QY 1367 TCTGCACCTGATGCGCGCTGTTTAAATGCTTGAAGCAAAAGCGACGCGGTATGCGGT 1426
Db 171 TCTCTGGCTGTGCTGCTT-----CTACCGTTCGCCACGCCGCGAGCGGTACGACAT 224
QY 1427 CAGCAGCTACCGCGATGTCATTCGCGCACTGGGCAATAGCGGACTTTCGCGGAAGTCA 1486
Db 225 CGCCGAGTACCGTGGCTGTCACAGGACTATCGCGAGCCTCGCGAGCTCGCGCGGTTCAT 284
QY 1487 TCTCTGGCTGTCAGAGCGCGCATTTTCGCGCTCTGTCGATTTTATCTTCAACCAACACTTC 1546
Db 285 CGCGAGGCGCCATCGAGCGGTCTGGGTGATTAACCGAGCTGCTGATCAACCACTACCTC 344
QY 1547 CAAACGAACAGAAATGGGCGCAACGCTGCGCCGCCCGCGGACCGCTTTTCGACAAATTTCTA 1606
Db 345 CGACCAAGCATCTCTGTTTCATCGTCCGCGCACGCGAAGAGGATCGCGCGCGCGCA 404
QY 1607 CTATATTTTCCCGACCGCGGATGCCGACCAATACGACCGCACCTTCGCGGAATCTT 1666
Db 405 CTACTACGCTCTGGTGGAC---AGCGACGAGAAATACAGGGTACCGGATCATCTTCAT 461
QY 1667 CCGGACACGACACCGCGGCGCTTCTCGCAACTTGAAGACGACGCTGGGTGTGAGACAC 1726
Db 462 CGACACCGAGAGTCCAACTGACACTTGGGACCGCGGTAGCCCAACAGTACTACTGGCACCG 521
QY 1727 CTTCAATCTCTTCCAAATGGGACTTGAATTAACGACACCGGTGGTATTCGCGCAATGSC 1786
Db 522 CTTCTATTTCCACAGCGCGGACCTGAACCTTCGACAAACCGCGAGGTCTCGCGGAGGTGCT 581
QY 1787 GCGGAAATGCTTCTTCCCAACTTGGGCGTTCGACATCTCGCTGATGATGCGGTGCTG 1846
Db 582 CGGGGTGATGCGCTACTGCTGGACATGCGGTGCGAGGCTCGCGCTGCGAGCGATTC 641
QY 1847 CTTTATTTGAAACAAATGGGACAAAGCTGCGAAACCTTGGCGAGCGGACCGCCCTCAT 1906
Db 642 CTACCTGATCGAACCGGACCGGACCAAGCAGCAGGAGAACCTTGGCGAGACCGCACGAGT 701
QY 1907 SCGCGGTTCAATGCGGTTATGCTGTTTCCGCGCGCGCGCTGTTCTTCAAAATCCGAAGC 1966
Db 702 CAAGCGCATCCGCGCGAGCTGACCGGCACTATTCGCGACCGCATGCTGCTGCGCGAGGC 761
QY 1967 CATGCTCCACCGGACCAAGTCTGTCCTCAATACATCGG-----GCAGGA 2008
Db 762 CAATCAGTGGCGCGGAGACACCGCGCGTACTTTCGCGCGGAGGATGGCGCGAGGCGGA 821
QY 2009 CGAATGCGAAATCGGTTACAACCCCTGCAATGGCATTTGTTGTGAACACCCCTTGCAC 2068
Db 822 CGAATGCGCAATGGCTTCCACTTTCGCGTGGTGGCGGATGATGCGGATGATGCGGATGCGCCA 881
QY 2069 GCGCGAAGTCAACCTGCTCCATCAGGCGGTGACCTACCGGCAACACTGCTCCGAGCATAC 2128
Db 882 GGAGGATCGTATCCGATCACCGACATCTGCGCCAGACCGCGGACATCCCGGCAATTG 941
QY 2129 CGCTGCGTCAACTAGTTCGCGAGCAGCAGCATCGGTGCGACGCTTTCGCGATGAAGA 2188
Db 942 CCAATGGCGGATCTTCTGCGCAACACGACGAGCTGACCTGGAGATGTTTACCGACGA 1001
QY 2189 CGCGGATATCTGGGATAGCGGCTTACGACCAACCGGCAATTTCTCAACCGCTTTCGT 2248
Db 1002 CGAGCGGACTATCTCTGGACCACTATGCGCGGCGCGCGGCGCTCAACCTGGG 1061
QY 2249 CAACCG 2254
Db 1062 CATCCG 1067

```

RESULT 7





Db 3442077 CGGCTTCTTCTCCACAGCCGACCTCAACTACGAGAACCGGGGCTGCAGGAGAGAT 3442136  
QY 1784 GGGGGGGAATGCTGCTTCTCCCAACTTGGCGGTGACATCTCGGTATGATGCGGT 1843  
Db 3442137 CGTCTCGGCTTGGGCTTCTGGCTGACCTCGCATCGAGGCTTCGGCTCGACGCGT 3442196  
QY 1844 TGCTTTTATTTGAAACAAATGGGCAAGCTGCGAAACCTGCGGACGCGCCCT 1903  
Db 3442197 GCGTACCTGTACACAGAGGACCACTGCGAGAACCTGCGGCGACGACGAGTT 3442256  
QY 1904 CATCCGGGCTTCAATGCGGTATGCGGTATGCGGCGCGCGGTCTTCTCAATCGA 1963  
Db 3442257 CTTGAAGCGGCTGCGAAGGAGATGACACGCACTATCCGACACGCTGCTGCGGA 3442316  
QY 1964 AGCATCTGCTCCACCCGACCAAGCTGCTCAATACATCGG 2002  
Db 3442317 GCGAACCAGTGGCGGAGGAGCTGCTGCTACTTTCGG 3442355

## RESULT 9

US-09-974-300-399/c

; Sequence 399, Application US/09974300

; Patent No. US20020146721A1

; GENERAL INFORMATION:

; APPLICANT: Berka, Randy M.

; TITLE OF INVENTION: Methods For Monitoring Multiple Gene

; TITLE OF INVENTION: Expression

; FILE REFERENCE: 10085.500-US

; CURRENT APPLICATION NUMBER: US/09/974,300

; CURRENT FILING DATE: 2001-10-05

; PRIOR APPLICATION NUMBER: 09/680,598

; PRIOR FILING DATE: 2000-10-06

; PRIOR APPLICATION NUMBER: 60/279,526

; PRIOR FILING DATE: 2001-03-27

; NUMBER OF SEQ ID NOS: 8481

; SOFTWARE: Fast-Seq for Windows Version 4.0

; SEQ ID NO 399

; LENGTH: 483

; TYPE: DNA

; ORGANISM: Bacillus licheniformis

US-09-974-300-399

Query Match 4.0%; Score 118; DB 10; Length 483;

Best Local Similarity 57.0%; Pred. No. 1,6e-26;

Matches 236; Conservative 0; Mismatches 175; Indels 3; Gaps 1;

QY 1 GAGTTTTCGCTTCCGAAACCGAAGTGAAGTGGAGCCGCAACCTGTCGGCAAGCGG 60

Db 424 GAAGCTTCTTTTGGGCAACCGAAGTGAAGTGGAGCCGCAACCTGTCGGCAAGCGG 365

QY 61 CTGACCGCCCTTTTGGGCAACCGAAGTGAAGTGGAGCCGCAACCTGTCGGCAAGCGG 117

Db 364 CTGATGATACCTTTTGGGCAACCGAAGTGAAGTGGAGCCGCAACCTGTCGGCAAGCGG 305

QY 118 TTGCTTTGAGCGTGGGCAAGCAAGTGAAGTGGAGCCGCAACCTGTCGGCAAGCGG 177

Db 304 TTGCTTTGAGCGTGGGCAAGCAAGTGAAGTGGAGCCGCAACCTGTCGGCAAGCGG 245

QY 178 GCAATTTTGCAGATGTCGGGCGCGAGTCTTCCATCTGTTTTCAGCGGCAATGATTTCT 237

Db 244 GCGATTTTGCAGATGTCGGGCGCGAGTCTTCCATCTGTTTTCAGCGGCAATGATTTCT 185

QY 238 TCTTGGCGGCGTGGGCAAGCAAGTGAAGTGGAGCCGCAACCTGTCGGCAAGCGG 297

Db 184 TCTGAGCTGGTGTCTTTTCAAAATCATGATTTAGATATATCATCGGACATCTTATGACCA 125

QY 298 TGAGCATGCGGCGGCGGCGGAGTCTTCCATCTGTTTTCAGCGGCAATGATTTCT 357

Db 124 TGTGCTGTTTCCAGCAGCGCTTTCATGCTTCTTCTGCTGATGAACAGCTGATATCAT 65

QY 358 ATGTGGGCGGCGGCTTTCATGAGGAGTGGAGCCGCAACCTGTCGGCAAGCGG 411

Db 64 ACATCGGCAAGCCCTGTTTAAACACCGCTTCATTTAAGAAATAATATCGCG 11

## RESULT 10

US-09-738-626-2527

; Sequence 2527, Application US/09738626

; Publication No. US20020197605A1

; GENERAL INFORMATION:

; APPLICANT: NAKAGAWA, SATOSHI

; APPLICANT: MIZOGUCHI, HIROSHI

; APPLICANT: ANDO, SEIKO

; APPLICANT: HAYASHI, MIKIRO

; APPLICANT: OCHIAI, KEIKO

; APPLICANT: YOKOI, HARUHIKO

; APPLICANT: TATEISHI, NAOKO

; APPLICANT: SENOH, AKIHIRO

; APPLICANT: IKEDA, MASATO

; APPLICANT: OZAKI, AKIO

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-125

; CURRENT APPLICATION NUMBER: US/09/738,626

; CURRENT FILING DATE: 2000-12-18

; PRIOR APPLICATION NUMBER: JP 99/377484

; PRIOR FILING DATE: 1999-12-16

; PRIOR APPLICATION NUMBER: JP 00/159162

; PRIOR FILING DATE: 2000-04-07

; PRIOR APPLICATION NUMBER: JP 00/280988

; PRIOR FILING DATE: 2000-08-03

; NUMBER OF SEQ ID NOS: 7059

; SOFTWARE: Patentin ver. 3.0

; SEQ ID NO 2527

; LENGTH: 1794

; TYPE: DNA

; ORGANISM: Corynebacterium glutamicum

US-09-738-626-2527

Query Match 3.8%; Score 111.8; DB 10; Length 1794;

Best Local Similarity 48.3%; Pred. No. 3.1e-24;

Matches 422; Conservative 0; Mismatches 427; Indels 24; Gaps 3;

QY 1305 GCGGCGATTTCGAAGGCTTGAAGATAAAATTCCTTATTTTCAAGAGCTTGGTTGACT 1364

Db 220 GTGCGATCGTTGAAGGCTTGAAGATAAAATTCCTTATTTTCAAGAGCTTGGTTGACT 279

QY 1365 TATCTGCACCTGATGCGCTGTTTAAATGCCCTGAAGGCAAAAGCGAGCGGCTATGCG 1424

Db 280 TGCATTGGATCCACCGCTTTTATGATTCCTCC-----ACTGCGGACGCGGCTTACGAT 333

QY 1425 GTGACGAGCTACGCGATGTCATCGGCACATAGGCGACTTTCGCGAAGTC 1484

Db 334 ATCCGCACTTCCGTAATCTGCGGAAATTCGCGACCGTGCATGACTTCGTGGAATC 393

QY 1485 ATTGCTGCGCTGCAGAACGCGGCAATTCGCGGCTGCTGATTTTATCTTCAACGACAC 1544

Db 394 GTTGACCAACGCGGCGGCTGCTGATTCATCCGACTTGGTCAATGAATCAACAC 453

QY 1545 TCCAGCAACAGATGCGCGCAACGCTGCGCGGCGGCGGCGGCTTTTCGCAATTC 1604

Db 454 TCCAGCAACAGATGCGCGCAACGCTGCGCGGCGGCGGCGGCTTTTCGCAATTC 513

QY 1605 TACTATATTTTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1664

Db 514 TTTATGTTGGAGGATGATCCACCTGTACAGGAGCGGCGGCGGCGGCGGCGGCGG 567

QY 1665 TTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1724

Db 568 GTAGATACAGAGAAATTCGAATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 627

QY 1725 ACCTTCAATTCCTTCCATGGGACTTGAATTAACAGCAACCGGTTGGTATTTCCGCGCAATG 1784

Db 628 CGCTTTTCTCCACCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 687

Qy	1785	CGCGGCGAAATGCTGTTCTCTTGCACAACTTGGCGCTTGACATCTCTGGCTATNGATTCGGGTT	1844
Db	688	CTAGATGCTTGCGGTTTCTGGCTGGACCTGGGACTTGATGGTTTTCGACTAGATTCGCGTT	747
Qy	1845	GCCTTTATTTTGGAAACAAATGGGACAAAGCTGCGAAACCTTCGCGCAGGCGCAGCGCCTC	1904
Db	748	CCATTATCTTTTGAACCGGAAGGCACCAACGCGGAAAACCTCAAGAAGAACCCACGATTTC	807
Qy	1905	ATCGCGCGTTTCAATGCGGTTATGCGTATTCGCGCGCGCGCGGTTTCTTCAAAATCCGAA	1964
Db	808	CTCAAACTGTGTGCTCTGTCTATTGAGAGGAATACCGCGCCGAATCTCTCTCGCAGAA	867
Qy	1965	GCCTATGTCACCCCGGACCAAGTGGTCCAAATACATTCGGCA-----GGACGAA	2012
Db	868	GCCAAACAAATGCGCCCGCAAGATGTGGTCCGAATCTTCGGTGAAGAAACAAGGCGCATGAA	927
Qy	2013	TGCCAAATCGGTTACAAACCCCGCTGCCAAATGGCAATGTTGTGGAACACCCCTTGCCACGCGC	2072
Db	928	TGCCACATGCGCTTCGACTTCCCTTTGATGCGCGGCACTCTCATGGGAGTTGCGCAAGGT	987
Qy	2073	GAAGTCAACCTGTCTCCATCAGCGGCTGACCTTACCGGCACCAACCTGCCCGGAGCATACGCGC	2132
Db	988	TACGCAACCCCGATCAGTGAATCTTGGCCAAACACCCCGGAGATTCCCAAGACTGCCCAA	1047
Qy	2133	TGGGTCAACTACGTTCGCGAGCCACGACGACATC	2165
Db	1048	TGGGGTATTTTCTGGGTAAATCATGATGAGCTC	1080

```

RESULT 11
US-09-738-626--i
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCES: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626--1

```

Qy	1425	GTGAGCAGCTACGGCGATGTCAAATCCGGCACTTGGGCACAATAAGGCGCACTTGGCGGAAGTCT	1484
Db	2438446	ATCCGCAACTTCCTCGTGAATACTCTGCCGAATTCGGGCACCGTCGATGACTTCTGTGGAATCTC	2438505
Qy	1485	ATTGCTCGCTGCACGAAGCGGCATTTCCGGCGTCTCGATTTCATTTATCTTCAACACACCC	1544
Db	2438506	GTTGACCAAGCCCAACCGCGTGGCTCGTGTTATCACCGACTTGGTCAATGAAATCACACC	2438565
Qy	1545	TCCAAAGAACAGAAATGGGCGCAAGCTGGCGCGCGCGACCCGCTTTTTCGACAAATTC	1604
Db	2438566	TCCGAACAGCAGCATGGTTCCAAGAAATCCGCGCGCAACCCAAACGGGCCCTTACGGAGAT	2438625
Qy	1605	TACTATATTTTCCCGACCGCGGATGCCGGAACCAATACGACGCGACCTTGGCGCGAAATC	1664
Db	2438626	TTCTATGTGTGGAGCGATGATCCACCTGTGTCAACGAAGCCGCAATCAT-----CTTT	2438679
Qy	1665	TTCCCCGACGACACCCCGGCGGCTTCTCGCAACTTGAAGACGGAGCGTGGGTGTGGACG	1724
Db	2438680	GTAGATACAGAGAATCCCAACTGGACCTATGATCCGGTGCCTGACGTACTTCTGGCAC	2438739
Qy	1725	ACCTTCAATTCCTTCCATGGGACCTTGAATTAACAGCAACCCGTGGGTATTCGGCGCAATG	1784
Db	2438740	CGCTTCTTCTCCACCAACGACACTCAACTACGCAACCCCGCGAGTCCAAAGAGGCCATG	2438799
Qy	1785	GCGGGCGAAATGCTGTTCCTTGCACATTGGCGCTTGACATCCTGCGTATGATGCGGTT	1844
Db	2438800	CTAGATGTCTTGGCTTCTGGCTGACCTTGGAACTTGATGGTTTCGACTAGATGCCGTT	2438859
Qy	1845	GCCTTTATTGGAAACAAATGGGCAACAGCTGCGAAACCTGCCGAGGCGCAGCGCCTC	1904
Db	2438860	CCTTATCTTTTGAACGCGAAGSCACCAACGGCGAAACCTCAAGAAACCCACGATTTTC	2438919
Qy	1905	ATCCGCGGCTTCAATGCGCTTATGCTATTGGCGCGCCGCGCTGTTCTTCAAAATCCGAA	1964
Db	2438920	CTCAAACTGTGCGCTCTCTCATTGAGAAGGAATACCCCGGCGGAATCTCTGCTCGAGAA	2438979
Qy	1965	GCCATCGTCCACCCGACCAAGTCGTCCAAATACATCGGSCA-----GGACGAA	2012
Db	2438980	GCCAACTAATGGCCCAAGATGTGTGGAATACTTCCGTGAAAAAGACAAGGCGATGAA	2439039
Qy	2013	TGCAAAATCGTTTAAACCCCTGCAATGGCAATTTGTTGTGAAACACCTTGGCAGCGC	2072
Db	2439040	TGCAATGCGCTTCCACTTCCCTTTGATGCGCGCATCTTCATGGGAGTTGCCAAGGT	2439099
Qy	2073	GAAGTCAACTGCTCCATCAGCGGTGACTACCGCCCAACCTGCCGAGCATACCGCC	2132
Db	2439100	TACGCAACCCGATCAGTGAGATCTTGGCCACACCCCGGAGTTCCTCCAGACTGCCCA	2439159
Qy	2133	TGGGTCAACTAGCTCGCAGCGCAACGACGACATC	2165
Db	2439160	TGGGGTAATTTCTCGGTAAATCATGATGCTC	2439192

```

RESULT 12
US-10-156-761-7374
Sequence 7374, Application US/10156761
Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHITUKU
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697

```

Qy	1992	CAATACATCGGGCAGGACGAATGCCAAATCGGTTACAAACCCCTGCAAAATGGCATTTGTG	2051
Db	799	CCGACACCGAACGTTGCGCGCTATCTGCGCCCGACGAACATGCACACGGCTTCAAC	858
Qy	2052	TGGAACACCTTCCACGCGGGGGAATCAACCTGCTCCATCAGGCGGTGACCTAC	2105
Db	859	TTCTCTCTCTCTCTCTGCGCCCTGGACGCGCGCGGCTTGGCAGCGGCATCGAC	912

## RESULT 14

```

US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIEBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIVUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (41877:5)
; OTHER INFORMATION: a, t, c, g, other or v
US-10-156-761-1

```

Query Match 2.4%; Score 70.8; DB 14; Length 9025608;  
Best Local Similarity 47.1%; Pred. No. 4.8e-09;  
Matches 336; Conservative 0; Mismatches 357; Indels 21; G

QY	1410	GACGGCGGCTATCGGTCAGCAGCTACCGCGATGTCAATCCGGCACTGGGCACAATAGGC	1469
Db	2106614	GACGGCGGCTATCGACGTGCGCGATACCGGGCCGTGAGCCCGGCTTCGGACACCTCGCC	2106555
QY	1470	GACTTGGCGAAGTCATTTGCTGCGCTGCAGGAAGCGGCATTTCCGCGCTGCTCGATTTT	1529
Db	2106554	GAAGCGGAGAAACTCATTCGCCGAGGCACGGGAACCTGGCATCCGCACCCCTTGTTCGACATC	2106495
QY	1530	ATCTTCAACCACTCTCAACGACACGAGATGGCGCGACGCTGCGCGCGCGCGGACCCG	1589
Db	2106494	GTCCCGAAACATGTCGCGACACGACCCCTGTTTCGCGCGCGCTGCGCGCGCGCCG	2106435
QY	1590	CTTTTCGACAAATTTCTACTATATTTTCCCGACGCGCGGATGCCGACCAATAACGACCGC	1649
Db	2106434	GGCAGCCCCGA---GGCGGCACGTTCCTCACTTCGCGGGGCGCGCGCACGCGGAA	2106378
QY	1650	ACCTGGCGGAATCTTCCCGACACGACACCGGG-----CGGCTTCTCGAA	1697
Db	2106377	CTCCCGCCCAACGACTGGCGGTCCCTAGTTTCGCGGCTCCACCGAACCCGTCTGACACAG	2106318
QY	1698	CTGGAAGAGGAGCGCTGGGTGTGCAGACCTTCAATTCCTTCCAAATGGGACTTGAATTAC	1757
Db	2106317	CTCCCCGACGGGACTGTGTACCTGCACCTGTTTCACCCCGACACGCGGACCTCACTGG	2106258
QY	1758	AGCAACCGTGGGTATTCGCGCAATGGCGGGGAAATGCTGTTCTCTGCCAATTTGGGC	1817
Db	2106257	GCCCAACCGGCGCTCCGCGAGGAACACGAGGACGTACTCGCTTCTGGTTCGAACGGGGC	2106198
QY	1818	GTTGACATCTCGGTATGGATGCGGTGTGCTTTATTGTGAA-----ACAAATGGGGACA	1871
Db	2106197	GTGCGGGGTACGCACTCGCGGCGCTCTTCGCAAGGACCCGGACCTGCCGGAC	2106138
QY	1872	AGCTGGGAAACCTGCGCAGGGCGCAGCGCTCATTCGCGCGTTCAATGCGTTATGCGT	1931
Db	2106137	CTCGTGCAGGGCGCGCCCGCACCCGTTGTCGACCGCGGACGAACTCCACGACATCTAC	2106078
QY	1932	ATTGCGCGCGCCGCGCTGTTTCTTCAATCCGGAAGCCATGCTCCACCCCGGACCAAGTCGTC	1991

Search completed: November 8, 2003, 20:14:57  
Job time : 917 secs

Db	2106077	CGTCTCGGCGCCCATCGCCGACGAGTACGGCGCGCTCTCTCGCGCAGGTCCTGGCTC	2106018
Qy	1992	CAATACATCGGCGAGCAGGAATGCCAAATCGGTTACAAACCCCTGCAAAATGGCATTTG	2051
Db	2106017	CCGACACGACGCTTGCGCCGCTATCTGCGCCCGAGAACTGACACCGCCTCAAC	2105958
Qy	2052	TGGAACACCTTGGCAGCGCGGAGTCAACCTCTCCATCAGGCGCTGACCTAC	2105
Db	2105957	TTCTCCTTCTCTCTGCCCCGCGGCGCTGCGCACGGCCATCGAC	2105904

RESULT 15

US-09-974-300-657

; Sequence 657, Application US/09974300

; Patent No. US20020146721A1

; GENERAL INFORMATION:

; APPLICANT: Beika, Randy M.

; TITLE OF INVENTION: Methods For Monitoring Multiple Gene

; TITLE OF INVENTION: Expression

; FILE REFERENCE: 10085.500-US

; CURRENT APPLICATION NUMBER: US/09/974,300

; CURRENT FILING DATE: 2001-10-05

; PRIOR APPLICATION NUMBER: 09/680,598

; PRIOR FILING DATE: 2000-10-06

; PRIOR APPLICATION NUMBER: 60/279,526

; PRIOR FILING DATE: 2001-03-27

; NUMBER OF SEQ ID NOS: 8481

; SOFTWARE: FastSeq for Windows version 4.0

; SEQ ID NO 657

; LENGTH: 1689

; TYPE: DNA

; ORGANISM: Bacillus licheniformis

US-09-974-300-657

Query Match	2.2%;	Score 65;	DB 10;	Length 1689;	
Best Local Similarity	52.7%;	Pred. No. 2.5e-09;			
Matches 168;	Conservative	0;	Mismatches 145;	Indels 6;	Gaps 1;
QY	1304	TGCGGCGATTGAAGGGCTTGAAGTAAATTCCTTATTTTCAGAGCTTCGGTTTGAC	1363		
Db	87	TGTGCGGCATACGCGGCATCATCGAAAGCTCGACTACATCAAGAGCTCGCCTGTGA	146		
QY	1364	TTATCTGCACCTGATGCGCGCTGTTTAAATGCCCTGAAGGCAAAAGCGACGGCGGCTATGC	1423		
Db	147	TGTCATTTGCTGACGCGGCTTATCATCGCC-----CCAAAACGACAACGGCTATGA	200		
QY	1424	GGTCAGCGCTACCGGAGTGTCAATCGGCATCGGGCAACAATAGGAGCTTTCGCGGAAGT	1483		
Db	201	CATCAGCGACTATTACAGCATCCATGAAGAAATATGGAAACAATGCCGACTTTGAAGAGCT	260		
QY	1484	CATTGCTGGCTGCACGAGCGCGCATTTCCGCGGTGCTGCATTTTATCTTCAACACAC	1543		
Db	261	TCTTGAAGAGGCCCAACAGCGCGGCATAAAGGTGATCATGGAATCTTGTTCGTCAACCATAC	320		
QY	1544	CTCCACGACACGAATTGGCGCAACGCTGCGCGCGCGGACCCGCTTTTCGACAAATTT	1603		
Db	321	ATCAACAGACCAAGATGGTTTAAAGAGCGGCGCTCCGCGCAAGAAAATCTGTACCCGGA	380		
QY	1604	CTACTATATTTTCCCGAC	1622		
Db	381	CTTTTATATTTTGGAAAGAC	399		

Search completed: November 8, 2003, 20:14:57  
Job time : 917 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 12:46:16 ; Search time 5415 Seconds

(without alignments)  
13079.085 Million cell updates/sec

Title: US-09-843-007A-1

Perfect score: 2914

Sequence: 1 GAGTTTCGGTCCCGAAC.....CGTTTCAGCGCATTTGCG 2914

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba.\*

2: em\_esthum.\*

3: em\_estin.\*

4: em\_estmu.\*

5: em\_estov.\*

6: em\_estpl.\*

7: em\_estro.\*

8: em\_hcc.\*

9: gb\_est1.\*

10: gb\_est2.\*

11: gb\_hcc.\*

12: gb\_est3.\*

13: gb\_est4.\*

14: gb\_est5.\*

15: em\_estfun.\*

16: em\_estom.\*

17: em\_gss\_hum.\*

18: em\_gss\_inv.\*

19: em\_gss\_pin.\*

20: em\_gss\_vrt.\*

21: em\_gss\_fun.\*

22: em\_gss\_mam.\*

23: em\_gss\_mus.\*

24: em\_gss\_pro.\*

25: em\_gss\_rtd.\*

26: em\_gss\_phg.\*

27: em\_gss\_vrl.\*

28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	104	3.6	1055	28	AF029476	AF029476
2	92.6	3.2	518	9	AI964123	AI964123
3	90.4	3.1	785	29	B2578463	B2578463
C 4	89.4	3.1	968	29	B2554240	B2554240

C 5	84.8	2.9	527	9	AI964122	AI964122
C 6	73.2	2.5	606	10	BE249225	BE249225
C 7	57	2.0	367	13	BU003396	BU003396
C 8	54.6	1.9	763	14	CB997861	CB997861
C 9	54.6	1.9	845	14	CB996585	CB996585
C 10	53	1.8	912	11	CNS08POT	CBX022785
C 11	50.4	1.7	1201	13	EX381961	EX381961
C 12	49.8	1.7	659	9	AI107032	AI107032
C 13	49.8	1.7	703	12	BMS87500	BMS87500
C 14	49.8	1.7	1981	28	BH770895	BH770895
C 15	49.2	1.7	551	9	AI386572	AI386572
C 16	49.2	1.7	589	9	AI517021	AI517021
C 17	49.2	1.7	589	12	BI366411	BI366411
C 18	49.2	1.7	603	12	BI372255	BI372255
C 19	49.2	1.7	604	12	BI288916	BI288916
C 20	49.2	1.7	607	12	BI579016	BI579016
C 21	49.2	1.7	654	12	BH173333	BH173333
C 22	49	1.7	884	28	BH149112	BH149112
C 23	48.4	1.7	925	28	BH132691	BH132691
C 24	48.4	1.7	727	13	BQ110645	BQ110645
C 25	47.8	1.6	638	9	AI402332	AI402332
C 26	47.8	1.6	641	9	AI292849	AI292849
C 27	47.4	1.6	857	28	AZ530946	AZ530946
C 28	46.2	1.6	644	14	CA631367	CA631367
C 29	46	1.6	527	12	BI628794	BI628794
C 30	46	1.6	531	12	BI566079	BI566079
C 31	46	1.6	534	9	AA803956	AA803956
C 32	46	1.6	568	9	AI107511	AI107511
C 33	46	1.6	570	12	BI566486	BI566486
C 34	46	1.6	590	12	BI586057	BI586057
C 35	46	1.6	604	9	AI238307	AI238307
C 36	46	1.6	615	12	BI587736	BI587736
C 37	46	1.6	621	12	BI564072	BI564072
C 38	46	1.6	621	12	BI634521	BI634521
C 39	46	1.6	623	9	AI389773	AI389773
C 40	46	1.6	626	9	AI297229	AI297229
C 41	46	1.6	637	12	BI241727	BI241727
C 42	46	1.6	643	9	AI063787	AI063787
C 43	46	1.6	648	12	BI583466	BI583466
C 44	45.6	1.6	645	14	CA559056	CA559056
C 45	45.2	1.6	497	29	CC169594	CC169594

ALIGNMENTS

RESULT 1  
AF029476/c  
LOCUS AF029476 1055 bp DNA linear GSS 29-AUG-2000  
DEFINITION AF029476.1 Salmonella typhimurium LT2, Lambda DASH II Salmonella  
typhimurium genomic clone 472-13, genomic survey sequence.  
ACCESSION AF029476  
VERSION AF029476.1 GI:2571006  
KEYWORDS GSS.  
SOURCE Salmonella typhimurium  
ORGANISM Salmonella typhimurium  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Salmonella.  
REFERENCE 1 (bases 1 to 1055)  
Wong,R.M.-Y., Wong,K.K., Senson,N.R. and McClelland,M.  
AUTHORS Sample sequencing of a Salmonella typhimurium LT2 lambda library:  
TITLE comparison to the Escherichia coli K12 genome  
JOURNAL FEMS Microbiol. Lett. 173 (2), 411-423 (1999)  
MEDLINE 93243757  
PUBMED 10227170  
COMMENT Contact: McClelland M  
Molecular Biology  
Sidney Kimmel Cancer Center  
3059 Science Park Road, San Diego, CA 92121, USA  
Email: mclelland@lifsci.sdsu.edu  
Class: shotgun.  
Location/Qualifiers  
source  
1. .1055

```
/organism="Salmonella typhimurium"
/mol_type="genomic DNA"
/strain="LT2"
/db_xref="taxon:602"
/clone="472-T3"
/clone_lib="Salmonella typhimurium LT2, Lambda DASH II"
/notes="Vector: Lambda DASH II; sequenced using Li-Cor
sequencer"
BASE COUNT 277 a 253 c 250 g 258 t 17 others
ORIGIN
Query Match 3.8%; Score 104; DB 28; Length 1055;
Best Local Similarity 58.3%; Pred. No. 1.5e-17;
Matches 201; Conservative 0; Mismatches 141; Indels 3; Gaps 1;
QY 1 GAGTTTGGCTTCCCGAAGCGAAGCGTGTGCTGAGCGCAACACCTGTCCGGCAAGCGG 60
DB 379 GATGCTTTTTCACCGCGCCAAACGTTGCGGAGAACCGACACTTCGCGCGGACAGCA 320
QY 61 CTGACCGCCCTTTTGGCCCATCGACATCGTAAACATCGTTTGTGGCAA---GCTCT 117
DB 319 GATATTACCCCGTTTTCGACATCGACATGTAATTAATCGGACATNCGATAGCGCTCC 260
QY 118 TTGCGTTTCAGCGTGCAGAAACCAAGTCAGACGCTTCCGCGTTTGGCGCATCAC 177
DB 259 TGCATTTCTACGTGGCGGTAAATGAGGTGAGACATCGGCTTTAGTCTGTGGCATGCG 200
QY 178 GCAATTTTGCAGATGTCGCGCGCGAGTCTCTCATCTGTTTCAGACGGCATACGATTTCT 237
DB 199 GGCATCTTCGGAATATCAGCGCCAGTCTCGCATTTTACGACAGACGCTGMAACAATCTCT 140
QY 238 TCTTGGCGGGCGTGGTGAATCATGATTCGACAGCAGGCGGCGGATCGCGTTT 297
DB 139 TCCGCTGGCGGGCTTTTATGAAATCATGTTAGACATGATCATCACCGCAATGTGTGA 80
QY 298 TGAGCATGCGCCACGCGCGCGCGCGCGGCTTTCGCGGAAAAA 342
DB 79 TGAGCATGCGCGCGCGCGCGCTTTTCACTCATCGTCCGCGTAAAA 35
RESULT 2
LOCUS AI964123 518 bp mRNA linear EST 20-AUG-1999
DEFINITION B2769237 zebrafish, Steve Ekker Danio rerio cDNA clone ZBAA17,
mRNA sequence.
ACCESSION AI964123 GI:5756836
VERSION AI964123.1
KEYWORDS EST.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.
AUTHORS Quackenbush, J., Lee, N., Tomb, J.-P. and Ekker, S.
TITLE Generation of ESTs from zebrafish
JOURNAL Unpublished
COMMENT Other ESTs: B2769236
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org.
FEATURES
source
1..518
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="ZBAA17"
/tissue_type="whole body"
/dev_stage="maternal"
/lab_host="XLI-Blue MRF"
```

```
/clone lib="zebrafish, Steve Ekker"
/notes="Vector: pAD-GAL4; Site 1: EcoRI; Site 2: XhoI; This
library was constructed by directed cloning into the
EcoRI-XhoI sites of pAD-GAL4. Estimated Average insert size
approximately 1.2kb based on 15 clones, estimated titer of
excised phagemid library = 3.78 x 10^6 cfu/ml"
BASE COUNT 127 a 121 c 123 g 141 t
ORIGIN
Query Match 3.2%; Score 92.6; DB 9; Length 518;
Best Local Similarity 56.4%; Pred. No. 1.6e-14;
Matches 189; Conservative 0; Mismatches 143; Indels 3; Gaps 1;
QY 1 GAGTTTGGCTTCCCGAAGCGAAGCGTGTGCTGAGCGCAACACCTGTCCGGCAAGCGG 60
DB 133 GACGCTTTTTCACCGCACCAAAAGTTCCCGCGGACCAATATCTTCCACCGCCAGACGA 192
QY 61 CTGACCGCCCTTTTGGCCCATCGACATCGTAAACATCGTTTGTGGCAA---GCTCT 117
DB 193 GAATATAGCCAGTTTTCCTCATCGACATCGTATATGACGATCGGCATCTGCTCC 252
QY 118 TTGCGTTTTCAGCGTGGCAGAAAGCAAGTCAGACGCTTCCGCGCTTTCGCGCATCAC 177
DB 253 TGCATCTCCAGGTTCGCGGCAAGCAACGTCAGACATCGCTGCTACTTTTGGCGCATCAGC 312
QY 178 GCAATTTTGCAGATGTCGCGCGCGAGTCTCTCATCTGTTTCAGACGGCATACGATTTCT 237
DB 313 GCAATCTTAGGAATATCGCGTTCGAAGGATTCATNTTTCGCGCAAGCAATGATTTCT 372
QY 238 TCTTGGCGGGCGTGGTGAACCTCATGATTCAGACAGCGCGCGCATGCGCTTTT 297
DB 373 TCGGCTTCCGCGTTTATGGAAGTCATGTTGGACATGACTTTCACATCATCGCGG 432
QY 298 TGAGCATGCGCCACGCGCGCGCGCGCGGCTTTC 332
DB 433 TGGCGGTAGGCAACGCGCTCANNCCGCTCGACT 467
RESULT 3
LOCUS BZ578463 785 bp DNA linear GSS 17-DEC-2002
DEFINITION msh2_5885.x1 msh Pseudomonas aeruginosa genomic clone msh2_5885,
genomic survey sequence.
ACCESSION BZ578463
VERSION BZ578463.1 GI:27213524
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 785)
AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE Whole-genome sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol., (2002) in press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES
Location/Qualifiers
1..785
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="MSH"
/db_xref="taxon:287"
/clone="msh2_5885"
/dev_stage="maternal"
/lab_host="XLI-Blue MRF"
/note="Environmental isolate. Whole genomic shotgun"
```

```
BASE COUNT      154 a      274 c      224 g      131 t      2 others
ORIGIN
Query Match      3.1%; Score 90.4; DB 29; Length 785;
Best Local Similarity 49.3%; Pred. No. 8.1e-14;
Matches 266; Conservative 0; Mismatches 271; Indels 3; Gaps 1;
Library."
QY 1476 CGGAAGTCATTCGTGGCTGTCACGAAGCCGGCATTTCCCGCGTCGTCGATTTTATCTTC 1535
Db 90 CGCGGTTTCATTCGCGGAGGCCATCGACGGGTCTGCGGGGATTTACCGAGCTGGTGATC 149
QY 1536 AACGACACTCCACGACGACGAATGGCGGCAAGCTGCGCGCGCGCGGACCCCTTTTC 1595
Db 150 AACGACACTCCGACGACGATCCCTGGTTTCATCCGCGCGCGCGGACCCGAGGATCG 209
QY 1596 GACAAATTTCTACTATATTTTCCCGGACCGCGGATGCCCAATACGACCGCACCTG 1655
Db 210 CGCGCCCGGACTA--CTAGCTTGGTGGACAGCGAGGAATACCGAGCTACCGG 266
QY 1656 CGGGAATCTTCCCGGACGACGACCGGGGGCTTTCTCGCAATCGGAAGACGACGCTGG 1715
Db 267 ATCATCTTTCATCGACACCGGACGATCCAACTGGACCTGGGACCGGTCGCCCAACAGTAC 326
QY 1716 GTGTGGACGACCTTCAATTCCTTCCATGGCACTTGAATTACGACACCGCTGGTATTC 1775
Db 327 TACTGGACCGCTTCTATTCCACCGCGGACCTGAACTTTCGACACCGGAGGTCCG 386
QY 1776 CGGCAATGGCGGCGGAATCTGTTCTTCCCAACTTGGCGGCTTGACATCTCGGCTATG 1835
Db 387 CGGAGGTGTGCGGGGTGATGGCTACTGCTGGTGGACATGGGGGTTCGAGCGCTGGCGCTG 446
QY 1836 GATGCGGTTGCTTTATTTGGAAACAAATGGGACGACGCTCGGAAACCTGCGCGAGG 1895
Db 447 GAGCGATTCCTTACCTGATCGAACCGGACGACGACGAGGAGAACTTCCGCGGAGAC 506
QY 1896 CAGCGCTCATTCGCGGCTTCAATTCGCTTATGGTATTTGCGCGCGCGCGCTTCTTC 1955
Db 507 CACGAGTGTCTCAGCGCATTCGCGCGGAGCTGGAGCGGCACTATCCCGACCGCATCTG 566
QY 1956 AATTCGAGGACGTCGTCACCGCGACCAAGTGTGCAATGTCGAGGAGGACGAAATGC 2015
Db 567 CTGCGCGAGGCGCAACCACTAGTGGCGGAGACACCGCGCGCTACTTTCGCTGCGAGGATGC 626

RESULT 4
BZ554240/c
LOCUS      BZ554240      968 bp      DNA      linear      GSS 17-DEC-2002
DEFINITION      pacs1-60_4587.y2 pacs1-60 Pseudomonas aeruginosa genomic clone
ACCESSION      BZ554240
VERSION      BZ554240.1 GI:27161328
KEYWORDS      GSS.
SOURCE      Pseudomonas aeruginosa
ORGANISM      Pseudomonas aeruginosa
REFERENCE      1 (bases 1 to 968)
AUTHORS      Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
              Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE      Whole-Genome-Sequence variation among multiple isolates of
JOURNAL      Pseudomonas aeruginosa library
COMMENT      J. Bacteriol., (2002) In press
              Contact: Chris K. Raymond
              Genome Center
              University of Washington
              Box 352145, Seattle, WA 98105-2145, USA
              Tel: 2062216954
              Fax: 2066857244
              Email: craymond@u.washington.edu
              Class: shotgun.
              Location/Qualifiers
              source
                  1. .968

BASE COUNT      154 a      274 c      224 g      131 t      2 others
ORIGIN
Query Match      3.1%; Score 90.4; DB 29; Length 785;
Best Local Similarity 49.3%; Pred. No. 8.1e-14;
Matches 266; Conservative 0; Mismatches 271; Indels 3; Gaps 1;
Library."
QY 1476 CGGAAGTCATTCGTGGCTGTCACGAAGCCGGCATTTCCCGCGTCGTCGATTTTATCTTC 1535
Db 90 CGCGGTTTCATTCGCGGAGGCCATCGACGGGTCTGCGGGGATTTACCGAGCTGGTGATC 149
QY 1536 AACGACACTCCACGACGACGAATGGCGGCAAGCTGCGCGCGCGCGGACCCCTTTTC 1595
Db 150 AACGACACTCCGACGACGATCCCTGGTTTCATCCGCGCGCGCGGACCCGAGGATCG 209
QY 1596 GACAAATTTCTACTATATTTTCCCGGACCGCGGATGCCCAATACGACCGCACCTG 1655
Db 210 CGCGCCCGGACTA--CTAGCTTGGTGGACAGCGAGGAATACCGAGCTACCGG 266
QY 1656 CGGGAATCTTCCCGGACGACGACCGGGGGCTTTCTCGCAATCGGAAGACGACGCTGG 1715
Db 267 ATCATCTTTCATCGACACCGGACGATCCAACTGGACCTGGGACCGGTCGCCCAACAGTAC 326
QY 1716 GTGTGGACGACCTTCAATTCCTTCCATGGCACTTGAATTACGACACCGCTGGTATTC 1775
Db 327 TACTGGACCGCTTCTATTCCACCGCGGACCTGAACTTTCGACACCGGAGGTCCG 386
QY 1776 CGGCAATGGCGGCGGAATCTGTTCTTCCCAACTTGGCGGCTTGACATCTCGGCTATG 1835
Db 387 CGGAGGTGTGCGGGGTGATGGCTACTGCTGGTGGACATGGGGGTTCGAGCGCTGGCGCTG 446
QY 1836 GATGCGGTTGCTTTATTTGGAAACAAATGGGACGACGCTCGGAAACCTGCGCGAGG 1895
Db 447 GAGCGATTCCTTACCTGATCGAACCGGACGACGACGAGGAGAACTTCCGCGGAGAC 506
QY 1896 CAGCGCTCATTCGCGGCTTCAATTCGCTTATGGTATTTGCGCGCGCGCGCTTCTTC 1955
Db 507 CACGAGTGTCTCAGCGCATTCGCGCGGAGCTGGAGCGGCACTATCCCGACCGCATCTG 566
QY 1956 AATTCGAGGACGTCGTCACCGCGACCAAGTGTGCAATGTCGAGGAGGACGAAATGC 2015
Db 567 CTGCGCGAGGCGCAACCACTAGTGGCGGAGACACCGCGCGCTACTTTCGCTGCGAGGATGC 626

RESULT 5
AI964122/c
LOCUS      AI964122      527 bp      mRNA      linear      EST 20-AUG-1999
DEFINITION      zebrafish, Steve Ekker Danio rerio cDNA clone R2BAAL7,
ACCESSION      AI964122
VERSION      AI964122.1 GI:5756835
KEYWORDS      EST.
SOURCE      Danio rerio (zebrafish)
ORGANISM      Danio rerio
REFERENCE      1 (bases 1 to 527)
AUTHORS      Quackenbush,J., Lee,N., Tomb,J.-F. and Ekker,S.
TITLE      Generation of ESTs from zebrafish
JOURNAL      Unpublished
COMMENT      Other ESTs: EST269237
              Contact: John Quackenbush
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850, USA
              Tel: 301 838 3528
              Fax: 301 838 0208
              Email: john@tigr.org
              3 prime sequence.
              Location/Qualifiers
              source
                  1. .527
                  /organism="Danio rerio"
                  /mol_type="mRNA"
```

```
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pacs1-60_4587"
/clone_lib="pacs1-60"
/note="clinical isolate 1-60 Whole genomic shotgun
library."
BASE COUNT      183 a      285 c      272 g      228 t
ORIGIN
Query Match      3.1%; Score 89.4; DB 29; Length 968;
Best Local Similarity 51.4%; Pred. No. 1.7e-13;
Matches 207; Conservative 0; Mismatches 196; Indels 0; Gaps 0;
Library."
QY 1634 CGACCAATACGACCGGACCGGATTCGCGGAAATCTTCCCGGACCGACCGCGGCTTCTC 1693
Db 520 CGAGAAATACCGGATTCGCGGATTCATTCATCGACCGGACGCTCCAACTGGACCTG 461
QY 1694 GCACTGGAAGACGAGCGTGGGTGGACGACCTTCAATTCCTTCAATGGGACTTGAA 1753
Db 460 GGACCGGTCGCGCCATCAGTACTTCTGCGGCTTCTATTCACGACCGGACCTGAA 401
QY 1754 TTACACCAACCGGTCGATTCGCGCAATGGCGGGAATGCTGTTCTTCCCACTT 1813
Db 400 CTTGCTCAACCGGACGCTCTGCGGAGGTGCTGGGGTGATGCGCTACTGGCTGGACAT 341
QY 1814 GGGCGTTGACATCTGCTGATGATGCGGTTGCGCTTTATTTGGAAACAAATGGGACAAG 1873
Db 340 GGGCGTCGACGCGCTGCGGCTGTACGCGATTCCTACCTGATCGAACGCGTCCGACCCAG 281
QY 1874 CTGCGGAACCTGCGCGAGCGGCGGCGCTCATCCGCGGTTCAATGCGGTTATGCGTAT 1933
Db 280 CAGCGAGAACCTTGGCGGAGACCCACAGGTGCTCAAGCGCATCCGCGCGAGCTGGACGC 221
QY 1934 TGCGCGCGCGCGGCTGTTCTTCAATTCGGAAGCATGCTCCACCGCGACCAAGTCGTCCA 1993
Db 220 GCACATATCGACCGCATGCTGCTGTCGCGGACCCACGAGTGGCGGAGACACCGCGCT 161
QY 1994 ATACATCGGGGAGGAGCAATGCAATCGTTTACACACCCCTG 2036
Db 160 GTACTTTCGTCGGGAGCATGACGACGACGCGGAGCAATCCCTG 118

RESULT 6
AI964122/c
LOCUS      AI964122      527 bp      mRNA      linear      EST 20-AUG-1999
DEFINITION      zebrafish, Steve Ekker Danio rerio cDNA clone R2BAAL7,
ACCESSION      AI964122
VERSION      AI964122.1 GI:5756835
KEYWORDS      EST.
SOURCE      Danio rerio (zebrafish)
ORGANISM      Danio rerio
REFERENCE      1 (bases 1 to 527)
AUTHORS      Quackenbush,J., Lee,N., Tomb,J.-F. and Ekker,S.
TITLE      Generation of ESTs from zebrafish
JOURNAL      Unpublished
COMMENT      Other ESTs: EST269237
              Contact: John Quackenbush
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850, USA
              Tel: 301 838 3528
              Fax: 301 838 0208
              Email: john@tigr.org
              3 prime sequence.
              Location/Qualifiers
              source
                  1. .527
                  /organism="Danio rerio"
                  /mol_type="mRNA"
```



Seq primer: TCACACAGGAACACGCTATGAC.  
Location/Qualifiers  
1..606  
/organism="Medicago truncatula"  
/mol\_type="mRNA"  
/db\_xref="taxon:3880"  
/clone="NF024D06DT"  
/tissue\_type="Plantlets"  
/dev\_stage="Pooled timepoints"  
/clone\_lib="Drought"  
/note="Vector: Lambda Zap; Contains a mixture of entire plantlets harvested in a series of days-post-watering timepoints."

BASE COUNT 151 a 153 c 147 g 155 t  
ORIGIN

Query Match 2.5%; Score 73.2; DB 10; Length 606;  
Best Local Similarity 49.8%; Pred. No. 4.7e-09;  
Matches 217; Conservative 0; Mismatches 21; Indels 6; Gaps 1;

QY 1738 TCCATGGGACTGGAATATACAGCAACCGTGGTATTCGGGCAATGGCGGCGAAATGC 1797  
Db |||||  
QY 564 TCCAAATTGACCTGAATACCGTAGCCCTGAAGTGTCTGCGCATGGTGTATTTAC 505  
Db |||||  
QY 1798 TGTTCCTTGGCAACTTGGCGTGTGACATCTCGGTATGATGCGGTGCTTTATTGGA 1857  
Db |||||  
QY 504 TCTGTACTCTTGGCAAGTGTCTGAGTATGTCGGCTGGATGCCGTGCTTTATGTGA 445  
Db |||||  
QY 1858 AACAAATGGGCAACAGCTGCGAAACCTGCGCGACGCGCCCTCATTCGCGGTTC 1917  
Db |||||  
QY 444 AAGACCGGGAACAGCTGCATCCATCTGGAATAAACACATCTGATTATCAAACTGTAC 385  
Db |||||  
QY 1918 ATGCGGTATGCGTATGCGCGCGCGCGGTGTTCTTCAATCCGAGCATCGTCCACC 1977  
Db |||||  
QY 384 GGTGATTTATGATAACGTTGCGCGAGGTACAGTATCATTTACCGAGCAATGTTCCGC 325  
Db |||||  
QY 1978 CGGACCAAGTCTCCAAATACAT-----CGGCGAGGAGCAATGCCAAATCGGTTACAACC 2031  
Db |||||  
QY 324 ATAAAGACAACTTCTTACTTTGCGGAGGCGATGACGAGCACATATGTTGTACCACT 265  
Db |||||  
QY 2032 CCTCCAAATGGCATTTGTTGGAACACCTTGGCGACGCGGAGTCACTACCTGCTCCATC 2091  
Db |||||  
QY 264 TCTCGTGGCGCACTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 205  
Db |||||  
QY 2092 AGGCGTGGCTACCGCCACACACCTGCGCGAGCATACGCTCGGTGCTCACTACGTCGCGA 2151  
Db |||||  
QY 204 CGTGGCGCAAACTGACACTACCTTCCAGCAACACCACTGTTTAACTTCTCTGCT 145  
Db |||||  
QY 2152 GGCACGACGATCGG 2167  
Db |||||  
QY 144 CTCAGATGGCATCGG 129  
Db |||||

RESULT 7  
BU003396  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

BU003396 367 bp mRNA linear EST 22-AUG-2002  
QGG35B11.YG.ab1 OG\_EFGHJ lettuce serriola Lactuca sativa cDNA clone  
QGG35B11, mRNA sequence.  
BU003396  
EST.  
BU003396.1 GI:22437791  
Lactuca sativa  
Lactuca sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
asterids; campanulids; Asterales; Asteraceae; Cichorioideae;  
Cichorieae; Lactuca.  
1 (bases 1 to 367)  
Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,  
Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison  
P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,  
Church,S., Jackson,L. and Bradford,K.  
Lettuce and Sunflower ESTs from the Compositae Genome Project

/db\_xref="taxon:7955"  
/clone="RZBAAL7"  
/tissue\_type="whole body"  
/dev\_stage="maternal"  
/lab\_host="XLI-Blue MRF"  
/clone\_lib="zebrafish Steve Ekker"  
/note="Vector: PAD-GAL4; Site 1: EcoRI; Site 2: XhoI; This library was constructed by directed cloning into the EcoRI-XhoI sites of PAD-GAL4. Estimated Average insert size approximately 1.2kb based on 15 clones, estimated titer of excised phagemid library = 3.78 x 106 cfu/ml"  
BASE COUNT 144 a 127 c 118 g 136 t 2 others  
ORIGIN

Query Match 2.9%; Score 84.8; DB 9; Length 527;  
Best Local Similarity 58.5%; Pred. No. 2.5e-12;  
Matches 185; Conservative 0; Mismatches 127; Indels 4; Gaps 2;

QY 1 GAGTTTGGCTTCCGCAACGACGATGATGCTTGGCGCAACACCTGTTCGCGCAAGCGG 60  
Db |||||  
QY 317 GAGCTTTTACCGCACCAAGTTCGGCGGAGCCAAATACCTTCAACGCGCAGCA 258  
Db |||||  
QY 61 CTGACCGCCCTTTTGGCCCATCGACATCGTAAACATCGTTTGGTGGCAA---GCTCT 117  
Db |||||  
QY 257 GAAATACGCCAGTTTTTGGCATCGATCGTGTGATAATTGGACGATCGGCTACTGCTCC 198  
Db |||||  
QY 118 FTGCGTTGAGCGTGGCAGAAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 177  
Db |||||  
QY 197 TGCATCTCAGGTCGCGGCAACAGCTGACATCGCTGTACTTTGGCGCATCAGC 138  
Db |||||  
QY 178 GCAATTTGACAGTTCGCGCGCGAGTCTCCATCTGTTTCAGAGCGCATAGATTCT 237  
Db |||||  
QY 137 GCAATCTTAGAATATCGCGTGGAGGATTTGCAATTTTGGCAGAGCGGCAATGTTCT 78  
Db |||||  
QY 238 TCTTGGCGCGGTCGGTGAACTCATGATTCAGAGCAGGCGCGGCGATGCGGTTTTT 297  
Db |||||  
QY 77 TC-GCTTCCGCGCTTTTATGAAAGTCATGTTGACATGATCTACTTTCATCATCGCGG 19  
Db |||||  
QY 298 TGAGCATGCGCCACGG 313  
Db |||||  
QY 18 TGGCGGTAGGCGAGG 3

RESULT 6  
BE249225/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BE249225 606 bp mRNA linear EST 13-JUL-2000  
NF024D06DT.F1047 Drought Medicago truncatula cDNA clone NF024D06DT  
5', mRNA sequence.  
BE249225  
EST.  
BE249225.1 GI:9119284  
Medicago truncatula (barrel medic)  
Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
Medicago.  
1 (bases 1 to 606)  
Torrez-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,  
Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.  
Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
Medicago truncatula drought library  
Unpublished  
Contact: May GD  
Plant Biology Division  
The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
Tel: 580 221 7391  
Fax: 580 221 7380  
Email: gdmay@noble.org  
Medicago Genome Initiative accession: MGI:S:27134  
Insert Length: 606 Std Error: 0.00  
plate: 024 row: D column: 06



**AUTHORS** NIH-MGC <http://mgc.nci.nih.gov/>  
**TITLE** National Institutes of Health, Mammalian Gene Collection (MGC)  
**JOURNAL** Unpublished  
**COMMENT** Contact: Robert Strausberg, Ph.D.  
 Email: cgbbs-remail.nih.gov  
 Tissue Procurement: Dr. Stefan Hansson  
 cDNA Library Preparation: Michael J. Brownstein (NHECRI) with help  
 and advice from Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: NDAM345 row: h column: 02  
 High quality sequence stop: 610.

**FEATURES** source  
 Location/Qualifiers  
 1..845  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:30330889"  
 /tissue\_type="pre-eclampsic placenta"  
 /lab\_host="DH10B Tona"  
 /clone\_lib="NIH MGC 148"  
 /note="Organ: Placenta; Vector: pBluescriptR; Site 1:  
 all-XhoI; Site 2: BamI; Library is oligo-dT primed and  
 directionally cloned using primer  
 5'-TTTTTTTTTTTNN-3', size-selected for average insert  
 size 2.3 kb and normalized to ROT 5. This is a primary  
 library enriched for full-length clones and constructed  
 using the Cap-trapper method (Carninci, in preparation).  
 Library constructed by M. Brownstein (NIMH/NHRI,  
 National Institutes of Health). Note: this is a NIH\_MGC  
 Library."

**BASE COUNT** 200 a 209 c 265 g 171 t

**Query Match** 1.9%; Score 54.6; DB 14; Length 845;  
**Best Local Similarity** 49.6%; Pred. No. 0.00089;  
**Matches** 172; Conservative 0; Mismatches 169; Indels 6; Gaps 1;

QY 1308 GCGATTGAGGGCTTGAAGATAAAATTCCTATTTCACAGAGCTTGGTTGACTTAT 1367  
 |||||  
 DB 443 GCGACTTTAAAGGGCTGACGGGAAACTCGATTACTACTCAATGCTCGGGGTGATTCG 384  
 |||||

QY 1368 CTGCACCTGATCGCGTGTTTAATGCCCTGAGGCAAAAGCGCGGCTATGGGTC 1427  
 |||||  
 DB 383 CTATGCTTACCACCTTCTATGACTCTCTCT-CTCCATGACGAGGTTACGACATT 330  
 |||||

QY 1428 AGCAGCTACCGCATGTCATTCGGCACTGGGCACTAGGCGACTTGGCGAAGTCATT 1487  
 |||||  
 DB 329 CGTGACTATCGCTGGATCCGGGAGGAGCTGGGCACCAATTGAGGCTTTAAGGTTTCCCT 270  
 |||||

QY 1488 GCTGGCGTGACGAAGCGGCATTTCCGCGTGTGCTGATTTATCTTACACCACTCC 1547  
 |||||  
 DB 269 GAGCGAGCCCATGACCGTGGCCCTTCGGCTATCATCGACTTCGTCATGAACCACTCG 210  
 |||||

QY 1548 AACGACACGAATGGGGCAACGCTGCGCGCGCGGCGACCGGCTTTTCGACATTTCTAC 1607  
 |||||  
 DB 209 GATTCCTATCCGTTGTTCCAGTCTCTACGCGCGGATCCGATCGTCTCTAAGGTAATAC 150  
 |||||

QY 1608 TATATTTTCCCGACCGCGGATGTCGCGACCAATACGACCGCACCTT 1654  
 |||||  
 DB 149 TAGCTATGGTGGGACACTGACGAGGCGCTACTCCGATGCTCTGATCAT 103  
 |||||

**RESULT 10**  
**CNS08PQT** 912 bp mRNA linear HTC 07-JAN-2003  
**LOCUS** Single read from an extremity of a full-length cDNA clone made from  
**DEFINITION** Anopheles gambiae total adult females. 5-PRIME end of clone  
 PK0AA34PD06 of strain 6-9 of Anopheles gambiae (African malaria  
 mosquito).

**ACCESSION** BX022785  
**VERSION** BX022785.1 GI:27572005  
**KEYWORDS** HTC  
**SOURCE** Anopheles gambiae (African malaria mosquito)  
**ORGANISM** Anopheles gambiae  
 Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;  
 Anopheles.  
**REFERENCE** 1 (bases 1 to 912)  
**AUTHORS** Genoscope.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (06-JAN-2003) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)  
 - Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
**FEATURES** source  
 Location/Qualifiers  
 1..912  
 /organism="Anopheles gambiae"  
 /mol\_type="mRNA"  
 /strain="6-9"  
 /db\_xref="taxon:7165"  
 /clone="PK0AA34PD06"  
 /plasmid="pME188-PL"  
 /note="end : 5-PRIME"  
**BASE COUNT** 214 a 231 c 290 g 177 t

**Query Match** 1.8%; Score 53; DB 11; Length 912;  
**Best Local Similarity** 53.3%; Pred. No. 0.0026;  
**Matches** 138; Conservative 0; Mismatches 115; Indels 6; Gaps 1;

QY 1304 TCCCGCGATTGAGGGCTTGAAGATAAAATTCCTATTTCACAGAGCTTGGTTGAC 1363  
 |||||  
 DB 144 TGTGGCGATCTGACCGGCATTACGGAGAGCTGGACATTTGGCCGATCTTGGTGTGAC 203  
 |||||

QY 1364 TTAATCTGCACTGATGCGCTGTTTAATGCTTGAAGCAAAAGCGCGGCTATGC 1423  
 |||||  
 DB 204 TGGGTGTGGCTGAGCCCGGTGTTTAATCAACCATCCGATGCG- - - - -GACTTTGGGTACGA 257  
 |||||

QY 1424 GGTCAAGAGTACCGCGCATGTCATTCGGCATCTGGGCAATAGCGCATTCGCGCAAGT 1483  
 |||||  
 DB 258 TATTTGGGACTTCGCGATGTTGATCCCAATCTTTGGCACTATGCGCGATTTGGACCGTAT 317  
 |||||

QY 1484 CATCTCGCTGTCAGAGCGGCGATTCGCGCGTCTGCGATTTTCTTCAACACAC 1543  
 |||||  
 DB 318 GGTGAGAGGCAAGACGCTCGGTATTAGTGATTTCTGGACTTTGTGCGGACCATAC 377  
 |||||

QY 1544 CTCCAACCAACACGAATCG 1562  
 |||||  
 DB 378 GAGCAGCAGCAGCAGTGG 396  
 |||||

**RESULT 11**  
**LOCUS** BX381961 1201 bp mRNA linear EST 08-MAY-2003  
**DEFINITION** BX381961 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
 clone CS0DI072YF05 3-PRIME, mRNA sequence.  
**ACCESSION** BX381961  
**VERSION** BX381961.1 GI:30453007  
**KEYWORDS** EST.  
**SOURCE** Homo sapiens (human)  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE** 1 (bases 1 to 1201)  
**AUTHORS** Li, W.-B., Gruber, C., Jesse, J. and Polayes, D.  
**TITLE** Full-length cDNA libraries and normalization  
**JOURNAL** Unpublished  
**COMMENT** Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: sequef@genoscope.cns.fr, Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. Contact : Feng Liang Email : [fliang@lifetech.com](mailto:fliang@lifetech.com) URL :



QY 1487 TGTGCGCTGCACGAAGCCGCAATTCGCGCTGCTGATTTTATCTTCAACACACCTC 1546  
 Db 302 CGACACGGCTTTGAGCTGGGCAATCAAGTGTGTTCTGCACTTTGTCGGCAATCACGCTC 361  
 QY 1547 CAACGAACAGAAATGGGCGCAACGCTGCGCGCGCGGCGACCCGCTTTTTCACAAATTCYA 1506  
 Db 362 GGATCAGCATGAGTGGTTCAGAAATGCTGCGGCAAGGAGCGCGGTTTACAGGATTT--- 418  
 QY 1607 CTATATTTTCCCGACCGCGGATGCGCGCAATACGACCGCACCCCTGCGCGAAATCTT 1666  
 Db 419 CTATGCTGGACGATGCTGCTCGAGGAGATGGCACTCGGGTGCACCACTG 478  
 QY 1667 CCGGACGACGACCGCGCGGCTTCTCGCAACTGGAGACGACGCTGGGTGTGGAGAC 1726  
 Db 479 GCAATCGGTGTTTACGATCCGCTTGGGAGTGCGACGAAGTCTGTCGACGATTAACCT 538  
 QY 1727 CTTCAATCTTC-----CAATGGGACTGAAATACAGCAACCGGCTTATTCGCGC 1780  
 Db 539 GCACCACTTACCAGGAGACCGGACTTGAATATCGCAATCCCAAGTGGTTCAGGC 598  
 QY 1781 AATGCGGCGGAAATGCTTCTTGGCAACTTGGGCGTTGACATCTCGGTATGGAT 1838  
 Db 599 CATGGATGATGTGTTGCTTTCTGCTCAACAGGCGGTACTGTTTTCGATCGAT 656

RESULT 13  
 BM587500  
 LOCUS  
 DEFINITION 17000687321295 A.Gam.ad.cDNA, blood1 Anopheles gambiae cDNA clone  
 19600449698387 5', mRNA sequence.

ACCESSION BM587500  
 VERSION BM587500  
 KEYWORDS EST.  
 SOURCE Anopheles gambiae (African malaria mosquito)  
 ORGANISM Anopheles gambiae  
 Eukaryota; Metazoa; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.  
 REFERENCE 1 (bases 1 to 703)  
 AUTHORS Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L., Charlab, R., Collins, F.H., Venter, J.C. and Hoffman, S.D.  
 TITLE Celera Anopheles gambiae EST project  
 JOURNAL Unpublished  
 COMMENT Contact: Holt R.A.  
 Celera Genomics  
 45 W. Gude Dr., Rockville, MD 20850, USA  
 Tel: 2404533151  
 Fax: 2404534580  
 Email: HoltR@celera.com  
 Plate: NU01004P70 row: F column: 05  
 Seq primer: M13 Reverse.

FEATURES  
 source  
 1..703  
 /organism="Anopheles gambiae"  
 /mol\_type="mRNA"  
 /strain="RSP-ST (Reduced susc. to Permethrin - std. chromosome)"  
 /db\_xref="taxon:7165"  
 /clone="19600449698387"  
 /dev\_stage="Adult"  
 /lab\_host="DHI0b"  
 /clone\_lib="A.Gam.ad.cDNA.blood1"  
 /note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole adult mosquitoes (mixed sex) frozen on liquid nitrogen 24 hours after human blood feeding. cDNA inserts >500 bp cloned directionally into pSport 1. Not 1 site is 3'. Clones available through the Malaria Research and Reference Reagent Resource Center (www.malaria.mr4.org)"  
 BASE COUNT 155 a 178 c 228 g 142 t

Query Match 1.7%; Score 49.8; DB 12; Length 703;  
 Best Local Similarity 52.5%; Pred. No. 0.018;

Matches 136; Conservative 0; Mismatches 117; Indels 6; Gaps 1;  
 QY 1304 TCCCGCGATTTGAAGGCTTGAAGATAAATCTCTTATTTCAAGAGCTTGGTTGAC 1363  
 Db 147 TGTGGCGGATCTGACCGGCAATTCAGGAGAGCTGGAACATTTTGGCGATCTTGGTGTGAC 206  
 QY 1364 TTATCTGCACCTGATGCCGCTGTTTAAATGCCCTGAAGCAAAAGCGACGGCGGTATGC 1423  
 Db 207 TGGGGTGTGGCTGAGCCCGGTGTTTAAATCACCGATGGCG-----GACTTTGGGTACGA 280  
 QY 1424 GGTACGACCTACCGGATGTCATCCGCACTGGGCAATAGGCGACTTGGCGAAGT 1483  
 Db 261 TATTTGCGACTTCCCGGATGTGATCCCAATCTTTGGCACTATGGCCGATTTGACCGTAT 320  
 QY 1484 CATTTGCTGCGCTGCACGAAGCCGCAATTCGCCGCTGTCGATTTTATCTTCAACACAC 1543  
 Db 321 GTTGCAGAGGCAAGACGCTCGTATTAAAGGTGATTTGCACTTTTGCCGAATCATAC 380  
 QY 1544 CTCACAGACACGATGG 1562  
 Db 381 GAGTGACGAGCAGATGG 399

RESULT 14  
 BH770895/c  
 LOCUS  
 DEFINITION LLMcag529 MG1363 Random Sequence Tag Library Lactococcus lactis subsp. cremoris genomic, genomic survey sequence.

ACCESSION BH770895  
 VERSION BH770895.1  
 KEYWORDS GSS.  
 SOURCE Lactococcus lactis subsp. cremoris  
 ORGANISM Lactococcus lactis subsp. cremoris  
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
 REFERENCE 1 (bases 1 to 1981)  
 AUTHORS Bolotin, A., Ehrlich, S.D. and Sorokin, A.  
 TITLE Studies of genomes of dairy bacteria Lactococcus lactis  
 JOURNAL Sci. Aliments, (2002) In press  
 COMMENT Contact: Sorokin, A.  
 Genetique Microbiome  
 INRA  
 CRJ INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France  
 Tel: 33 1 34 65 25 15  
 Fax: 33 1 34 65 25 21  
 Email: sorokine@jouy.inra.fr  
 best homologue in strain IL1403 is amyV (90%)  
 Class: shotgun  
 High quality sequence start: 30  
 High quality sequence stop: 1953.  
 Location/Qualifiers  
 1..1981  
 /organism="Lactococcus lactis subsp. cremoris"  
 /mol\_type="genomic DNA"  
 /strain="MG1363"  
 /db\_xref="taxon:1359"  
 /clone\_lib="MG1363 Random Sequence Tag Library"  
 /note="Vector: pSGM2; Site 1: SmaI; Library of chromosomal fragments of L.lactis strain MG1363 was prepared by partial AluI digestion or by sonication."

BASE COUNT 574 a 429 c 335 g 643 t  
 ORIGIN

Query Match 1.7%; Score 49.8; DB 28; Length 1981;  
 Best Local Similarity 54.7%; Pred. No. 0.03;  
 Matches 99; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 1406 AAGCGAGCGCGGTATGCGGTACGAGCTACCGGATGTCAATCCCGCACTGGGCAAT 1465  
 Db 1634 AAGCTATCATGCTATGCTGACCAATACGAGAAATTAATCTTAATTTGGCAAT 1575  
 QY 1466 AAGCGACTTGGCGAAGTCAATTCCTCGCTGACGAAGCGGCAATTCGCGCTCGA 1525

Db 1574 GGCTGATTTTGAATACTGATTCCTCAGGCTATAAAACAGGAATTCCTGGATTCCTGATCACTCAAGT 1515  
 QY 1526 TTATTACTTCAACACACCTCCACACGACGATGGGCCACACGCTGCGCCGCGCGGA 1585  
 Db 1514 CATCGCGTTTATCCACACGCGGACTGACATGTCGTGGTTTCAAAAAGCAGTACTAGCTGGCGA 1455  
 QY 1586 C 1586  
 Db 1454 C 1454

## RESULT 15

AI386572 551 bp mRNA linear EST 19-APR-2001  
 GH16695.5prime GH Drosophila melanogaster head pOT2 Drosophila  
 melanogaster cDNA clone GH16695 5prime similar to V00204: Lvph  
 PBgn0002570 PID:g7777 SWISS-PROT:P07190, mRNA sequence.  
 AI386572  
 AI386572.1 GI:4200583  
 EST.  
 Drosophila melanogaster (fruit fly)  
 Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 551)  
 Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,  
 Lewis, S. and Rubin, G.M.  
 BDGP/HMI Drosophila EST Project  
 Unpublished  
 Contact: Stapleton, M.  
 BDGP  
 Lawrence Berkeley National Lab  
 One Cyclotron Rd, Berkeley, CA 94720, USA  
 Fax: 510 486 6798  
 Email: [http://www.fruitfly.org/EST\\_est@fruitfly.berkeley.edu](http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu)  
 hit genomic sequence AC005430; hit genomic sequence AC005451  
 Plate: 166 row: H column: 11  
 High quality sequence stop: 544.  
 Location/Qualifiers  
 1. .551

FEATURES  
source

/organism="Drosophila melanogaster"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7227"  
 /clone="GH16695"  
 /sex="male and female"  
 /dev\_stage="adult"  
 /lab\_host="DH5 - alpha"  
 /clone\_lib="GH Drosophila melanogaster head pOT2"  
 /note="Organ: head; Vector: pOT2; Site 1: EcoRI; Site 2:  
 XhoI; Sized fractionated cDNAs were directly ligated into  
 pOT2. Plasmid cDNA library."  
 BASE COUNT 139 a 152 c 150 g 110 t  
 ORIGIN

Query Match 1.7%; Score 49.2; DB 9; Length 551;  
 Best Local Similarity 50.3%; Pred. No. 0.024;  
 Matches 151; Conservative 0; Mismatches 143; Indels 6; Gaps 1;  
 QY 1308 GCGATTTGAGGGCTTGAAGATAAAATCTTATTTTCAGAGCTTGCTTTGACTTAT 1367  
 Db 128 GCGACCTGACGGGGTCACTGAAAAGCTGCAGTACTCTGAAAGACATCGGCTTCACGGGC 187  
 QY 1368 CTGCACCTCATCGCTGTTTAAATGCCCTGAAGGCAAAAGCGGGGCTATCGGTC 1427  
 Db 188 ACATGCTGTCGCCCATATTCAATCGCC-----CATGGTCGCTTTGGTTACACATA 241  
 QY 1428 AGCAGTACCGGATGTCATCCGGCACTGGGACAAATAGGCGACTTGGCGAAGTCAAT 1487  
 Db 242 TCGGACTTCTACAGATCCATCCGAATATGAAACCATGGAGACTTTGAGCGAATGATC 301  
 QY 1488 GTGGGCTGCAGAGCGGGCATTTCCGGCGCTGCTGATTTATCTTCAACCCAGCTCC 1547

Db 302 GCCAGGCCCAAGGAGGTGGGCATTAAATCATCTCTGGATTCCTGATCAACCACTCAAGT 361  
 QY 1548 AACGAACACGAATGGCGCGCAACGCTGCGCGCGCGGACCCGCTTTTCGACAAATTTCTAC 1607  
 Db 362 ACCGAAAACGAATGGTTTCAACAAGTCTGTGGACAGTACCCCGTCTCTACAAAGACTTCTAC 421

Search completed: November 8, 2003, 17:31:23  
 Job time : 5420 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 8, 2003, 17:35:13 ; Search time 5434 Seconds  
(without alignments)  
4788.098 Million cell updates/sec

Title: US-09-843-007A-2

Perfect score: 3408

Sequencing: 1 MLTPQQVGLIQLKTRIL.....VSLNQDLTLQYQVMWLETA 636

Scoring table:

BLOSUM62 Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame.p2n.model -DRV=xlp  
-Q/cgn2\_1/USPTO.spool\_p/US09843007/runat\_07112003\_170420\_25346/app\_query.fasta\_1.775  
-DB=GenEmbl -QFWT=fastap -SUPFIX=p2n.rge -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pco -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09843007@cgn\_1\_1\_4958@runat\_07112003\_170420\_25346 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -JONLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.ov.\*  
5: gb.pat.\*  
6: gb.ph.\*  
7: gb.pl.\*  
8: gb.pr.\*  
9: gb.ro.\*  
10: gb.sts.\*  
11: gb.sy.\*  
12: gb.un.\*  
13: gb.vi.\*  
14: gb.da.\*  
15: em.fun.\*  
16: em.hun.\*  
17: em.in.\*  
18: em.mi.\*  
19: em.ro.\*  
20: em.ov.\*  
21: em.or.\*  
22: em.pat.\*  
23: em.ph.\*  
24: em.pl.\*  
25: em.ro.\*  
26: em.sts.\*  
27: em.un.\*

29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sv.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3408	100.0	1939	6	AX107013	Sequence
2	3408	100.0	2115	1	NPO011781	AXJ011781 Neisseria
3	3408	100.0	6878	6	AX107015	Sequence
4	3360	98.6	1910	6	AX107014	Sequence
5	3360	98.6	6851	6	AX107016	Sequence
6	3338	97.9	1967	1	AY099334	Neisseria
7	3338	97.8	1967	1	AY099335	Neisseria
8	3141.5	92.2	2883	6	A47369	Sequence 1
9	3141.5	92.2	2883	6	A47785	Sequence 1
10	1319.5	38.7	10719	1	AE001346	AE001946 Deinococcus
11	1126.5	33.1	10261	1	AE011997	Xanthomon
12	1116.5	32.8	11605	1	AE012454	Xanthomon
C 13	915.5	26.9	10822	1	AE005791	CAulobact
C 14	550	16.1	304517	1	AE016789	Pseudomon
C 15	548.5	16.1	189050	1	AL646077	Ralstonia
16	543.5	15.9	17863	1	AE004642	Pseudomon
17	529.5	15.5	1794	6	AX122611	Sequence
18	529.5	15.5	1794	6	BD164728	Nevel pol
19	529.5	15.5	1981	6	AX469814	Sequence
20	529.5	15.5	9184	1	TS0001206	Streptomy
21	529.5	15.5	281450	1	AP005032	Streptomy
C 22	529.5	15.5	303550	1	SC0939131	Streptomy
23	529.5	15.5	326150	1	AP005281	Corynebact
24	529.5	15.5	349980	6	AX127150	Sequence
25	529.5	15.5	349980	6	AX127151	Sequence
26	529	15.5	11866	1	AE012109	Xanthomon
27	526.5	15.4	306650	1	AP005221	Corynebact
C 28	525.5	15.4	10134	1	AE012958	Chlorobiu
29	523.5	15.4	12548	1	AE011639	Xanthomon
30	522	15.3	298600	1	AP005959	Bradyrhiz
31	521.5	15.3	309904	1	AE016865	Pseudomon
32	520.5	15.3	10890	1	AE002040	Deinococc
C 33	517.5	15.2	10718	1	AE010836	Methanosa
34	516	15.1	9589	1	SC0001205	Streptomy
C 35	515.5	15.1	300100	1	SC0939123	Streptomy
36	515	15.1	4121	1	AF135796	Thermus c
37	513	15.1	2895	6	BD180010	Highly th
38	511.5	15.0	325483	1	AP005050	Streptomy
39	509.5	15.0	11700	1	MTC14188	Mycobacteri
40	509.5	15.0	14029	1	AE006925	Mycobacteri
41	509.5	15.0	343050	1	BX248334	Mycobacteri
42	504	14.8	1704	6	AR012159	Sequence
43	504	14.8	1704	6	AR012160	Sequence
44	504	14.8	1704	6	E10495	gDNA encodi
45	504	14.8	2304	1	D78198	Pimelobacte

# ALIGNMENTS

RESULT 1



AX107013	AX107013	Sequence 4 from Patent WO0125449.	1939 bp	DNA	linear	PAT 30-APR-2001
LOCUS	AX107013	AX107013				
DEFINITION	AX107013	AX107013				
ACCESSION	AX107013	AX107013				
KEYWORDS	AX107013.1	GI:13922547				
SOURCE		synthetic construct				
ORGANISM		artificial sequences.				
REFERENCE						
AUTHORS		Bengs, H., Polakowski, T., Held, A. and Gallert, K.C.				
TITLE		Novel immobilizable amylase, the use thereof and method for				
		producing poly(1,4-alpha-glucan)				
JOURNAL		Patent: WO 0125449-A 4 12-APR-2001;				
		Axiva GmbH (DE)				
FEATURES						
source		Location/Qualifiers				
		1..1939				
		/organism="synthetic construct"				
		/mol_type="genomic DNA"				
		/db_xref="taxon:32630"				
		/note="PCR-Amplikon, kodierend fur amylase (Amsu)"				
BASE COUNT	453 a	605 c	482 g	395 t	4 others	
ORIGIN						
Alignment Scores:						
Pred. No.:	5,65e-287	Length:	1939			
Score:	3408.00	Matches:	636			
Percent Similarity:	100.00%	Conservative:	0			
Best Local Similarity:	100.00%	Mismatches:	0			
Query Match:	100.00%	Indels:	0			
DB:	5	Gaps:	0			
US-09-843-007A-2 (1-636) x AX107013 (1-1939)						
QY	1	MetLeuThrProThrGlnValGlyLeuLeuGlnTyrLeuLysThrArgGlyLeu	20			
DB	12	ATGTTGACCCCGCCAGAGTGGTTCATTTACAGTACCTCAAAACACGACATCTG	71			
QY	21	AspIleTyrThrProGluGlnArgAlaGlyIleGluLysSerGluAspThrArgGlnPhe	40			
DB	72	GACATCTACAGCGCCGACAGCGCGCGGATCGGAAATCCGAAAGACTGGCGGCAATTT	131			
QY	41	SerArgArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGly	60			
DB	132	TCGCGCGCGCATGGATAGCATTTCCCAACATGATGACGACCTCGACAGCGTGTACGGC	191			
QY	61	AsnAsnGluAlaLeuLeuProMetLeuGluMetLeuLeuAlaGlnAlaThrGlnSerTyr	80			
DB	192	AACAACGAAGCCCTGCTGCTATCTGGAATGCTGCTGGCGCAGCGCATGGCAAGCTAT	251			
QY	81	SerGlnArgSerSerLeuLysAspIleAspIleAlaArgGluAsnAsnProAspThr	100			
DB	252	TCCCAACGCAACTCATCTTAAAGATATCGATATCGCGCGGAAACACACCCCGATGG	311			
QY	101	IleLeuSerAsnLysGlnValGlyValCysTyrValAspLeuPheAlaGlyAspLeu	120			
DB	312	ATTTTGTCCAAACAAACAAAGTCGGCGGCGTGTGCTACGTTGATTTGTTGCGGCGAATTTG	371			
QY	121	LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeu	140			
DB	372	AAGCGCTTGAAGATAAAATCCCTATTTCAAGAGCTTGGTTGATTCATCTGCACCTG	431			
QY	141	MetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyr	160			
DB	432	ATGCGCTGTTTAAATGCCCTGAGGCAAAAGCGCGCGCTATGCGGTCAGCAGCTAC	491			
QY	161	ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaLeu	180			
DB	492	CGCGATGTCAATCCGCGACATGGGCACAAATAGCGCACTTGGCGAAGTCATTGCTGCGCTG	551			
QY	181	HisGluAlaGlyIleSerAlaValAspPheIlePheAsnHisThrSerAsnGluHis	200			
DB	552	CACGAAGCCGCAATTTCCGCGCTGCTCGATTTTCTTCACCAACACCTCCACGACAC	611			

QY	201	GluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePhe	220			
DB	612	GAATGGCGCAACGCTGCGCGCGCGCGCGCTTTTCGACAAATTTCTACTATATTTC	671			
QY	221	ProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGln	240			
DB	672	CCCGACCGCGGATGCGGACCAATACGACCGCACCTGCGGAAATCTTCCCGCACAG	731			
QY	241	HisProGlyGlyPheSerGlnLeuGluAspGlyArgTyrValTyrThrPheAsnSer	260			
DB	732	CACCGCGCGGCTTTCGCACTGGAGACGACGCTGGTGTGGACGACCTTCAATTC	791			
QY	261	PheGlnTyrAspLeuAsnTyrSerAsnProThrValPheArgAlaMetAlaGlyGluMet	280			
DB	792	TTCCAATGGCACTTCAATATACAGCAACCGCTGGTATTCCGCGCAATGCGCGCGAAATG	851			
QY	281	LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTyr	300			
DB	852	CTGTTCTTTCGCAACTTGGCGCTTGACATCTTCGCTATGGATGCGGTGCTTATTTGG	911			
QY	301	LysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPhe	320			
DB	912	AAACAAATGGGACCAAGCTGCGAAACCTGCGCGAGGCGACGCCCTCATCCGCGCTTC	971			
QY	321	AsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHis	340			
DB	972	AATGCGCTTATCGTATTCGCGCGCGCGCTGTTCTTCAATCCGAAACCTATCGTCCAC	1031			
QY	341	ProAspGlnValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnProLeu	360			
DB	1032	CCCGACCAAGTCTGTCATACATCGGCGAGGACGAATGCGCAATCGTGTACAAACCCCTG	1091			
QY	361	GlnMetAlaLeuLeuThrAsnThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAla	380			
DB	1092	CAATGGCACTTGTGTGAAACCTTGCACGCGCGAAAGTCAACCTGTCTCATCAGGGG	1151			
QY	381	LeuThrTyrArgHisAsnLeuProGluHisThrAlaTyrValAsnTyrValArgSerHis	400			
DB	1152	CTGACCTACCGCAACCTGCGCGACATACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1211			
QY	401	AspAspIleGlyTyrThrPheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyr	420			
DB	1212	GACGACATCGGCTGGACGTTTCGATGAGACGCGGCAATCTGGGCGATAAGCGGCTAC	1271			
QY	421	AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg	440			
DB	1272	GACACCGCGCAATTCCTCAACCGCTTCTTCGTCACCGCTTTCGACGCGAGCTTCGCTCGT	1331			
QY	441	GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla	460			
DB	1332	GGCGTACCGTTCCATATCAACCAAGCACAGCGACTGCGCTGCTAGTGGTACAGCGCGG	1391			
QY	461	AlaLeuValGlyLeuAlaGlnAspAspProHisAlaValAspArgIleLysLeuLeuTyr	480			
DB	1392	GCATTCGTCGCTTGGCGAAGCAGATCCCGCGCTTTCGCGCATCAAACTCTTGTAC	1451			
QY	481	SerIleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr	500			
DB	1452	AGCATTCGTTTGATACCGCGCTCTCGCGCTGATTTACCTAGGCGACCAAGTGGGTACG	1511			
QY	501	LeuAsnAspAspAspThrSerGlnAspSerAsnLysSerAspAspSerArgTyrAlaHis	520			
DB	1512	CTCAATGACGACGACTGGTTCGCAAGACACCAATAGAGGACGACAGCGCTTGGGCGCAC	1571			
QY	521	ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGly	540			
DB	1572	CGTCCCGGTACACGACGCTTGTACGCGCAACGCAACGATCCGTCGACCGCGCGCGG	1631			
QY	541	GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp	560			
DB	1632	CAAACTCTATCAGGCTTGGCGCATATGATTCGCGTCCGCGCAAGCAATCCGCGCTTCGAC	1691			

QY	561	GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArgAsn	580
Db	1692	GGCGCAGGCTGGTTACATTCACCAACCAACAGCACATCATCGGTACATCCGCAAC	1751
QY	581	AsnAlaLeuLeuAlaPheGlyAsnPheSerGlyThrProGlnThrValThrAlaHisThr	600
Db	1752	AATGGCGCTTTGGCATTCGGTAACCTTCAGCGAATATCCGCAACCGTTACCGCGCATACC	1811
QY	601	LeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyLysThrValSerLeuAsn	620
Db	1812	CTGCAAGCATGCCCTTCAGCGCGCAGACCTCATCGTGGCGCAAACTGTACGCTGAT	1871
QY	621	GlnAspLeuThrLeuGlnProTyrGlnValMetTrpLeuGluIleAla	636
Db	1872	CAGGATTTGACGCTTCAGCGCTCATCAGGTTCATGTGGCTCGAANTGCC	1919
RESULT 2			
NPO011781			
LOCUS	NPO011781	2115 bp	DNA linear BCT 28-JAN-1999
DEFINITION	Neisseria polysaccharea amylosucrase gene.		
ACCESSION	AJ011781		
VERSION	AJ011781.1	GI:4049491	
KEYWORDS	amylosucrase; amylosucrase gene.		
SOURCE	Neisseria polysaccharea		
ORGANISM	Neisseria polysaccharea		
REFERENCE	Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Neisseria.		
AUTHORS	De Montalk, G.P., Remaud-Simeon, M., Willemot, R.M., Planchot, V. and Monsan, P.		
TITLE	Sequence analysis of the gene encoding amylosucrase from Neisseria polysaccharea and characterization of the recombinant enzyme		
JOURNAL	J. Bacteriol. 181 (2), 375-381 (1999)		
MEDLINE	99102197		
PUBMED	9882648		
REFERENCE	2 (bases 1 to 2115)		
AUTHORS	Potocki de Montalk, G.		
TITLE	Direct Submision		
JOURNAL	Submitted (02-OCT-1998) Potocki de Montalk G., U.M.R. C.N.R.S. 5504, L.A. I.N.R.A. D.G.B.A., I.N.S.A., Centre de bioingenierie Gilbert Durand, Complexe scientifique de Rangueil, Toulouse cedex, 31077, FRANCE		
FEATURES			
source	Location/Qualifiers		
	1..2115		
	/organism="Neisseria polysaccharea"		
	/mol_type="genomic DNA"		
	/strain="ATCC 43768"		
	/db_xref="taxon:489"		
gene	136..2046		
	/gene="amylosucrase"		
CDS	<136..2046		
	/gene="amylosucrase"		
	/EC_number="2.4.1.4"		
	/codon_start=1		
	/transl_table=11		
	/product="amylosucrase"		
	/protein_id="CAA09772.1"		
	/db_xref="GI:4107260"		
	/translation="MLTPQQVGLILQVILKTLIDITYPEOPAGIEKSDKRPGRSM DTHFKLWELDSVGNELALPMELLLAQWQSYQNSSLKXIDIDARENPDWIL SNKQGVGVCTVDFAGLKLKDKIPIYQELGLTYLHLPLPKFEGSKSDGYVSSY RDVNFALGTGLRVLIALHEAGISAVVDIFNHTSHEHWAQCAAGDPLFDNFYI IPDRMPQYDTRLEIPDQHPGSPQSGLEDGRVWTFNFSQNDLNYSNPWFPM ACEMLFLVDILRMDAVFIWKQGTSCENLPQAHALIAFNANVRILAAVAVFK SFAIVHPQVQVIGQECQIGYNPLQWALLWNTLATREVLHLLHQALTYRHLPEHTA WNYVSRDDIGWTFADDAAVLGIISGVYDHPQPLNRFVNRFDGSAFSGVPQYNPST GDCRVSGTAAUVLGAQDDPHAVDRILKILYSIALSTGGLPLIYLGDVGLNDQDSQ DSKSDSDSWAHRPRYNEALYQARDPSTAAGQIYQGLNHLIAVQSNRPFQDGRILVT FNTNKKHIGIYRNALLAFGNFSEYPTQVTTAHTLQAMPFFKAHDILGKTKVSLNGLT LQPYQVMLEIA"		
BASE COUNT	493 a	529 g	432 t
ORIGIN			

Alignment Scores:			
Pred. No.:	6,39e-287	Length:	2115
Score:	3408.00	Matches:	636
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0
US-09-843-007A-2 (1-636) x NPO011781 (1-2115)			
QY	1	MetLeuThrProThrGlnGlnValGlyLeuIleLeuGlnTyrLeuLysThrArgIleLeu	20
Db	136	ATGTTGACCCCAACGACGACAGTGGTTTGATTTTACAGTACCTCAAAACAGCATCTTG	195
QY	21	AspIleTyrThrProGlnGlnArgAlaGlyIleGlyLysSerGluAspTrpArgGlnPhe	40
Db	196	GACATCTACACGCCCAACAGCGCGCGCATCGAAAAATCCGAAGATCGCGGCAGTTT	255
QY	41	SerArgArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGly	60
Db	256	TCGCGCGCATGATACGCAATTTCCCHACTGATGAACGAACTCGACAGCGGTGTACGCG	315
QY	61	AsnAsnGluAlaLeuLeuProMetLeuGluMetLeuLeuAlaGlnAlaTrpGlnSerTyr	80
Db	316	AACAACGAAGCCCTGCTGCTATGCTGAAATGCTGCTGGCGCAGGCGATGCAAGCTAT	375
QY	81	SerGlnArgAsnSerSerLeuLysAspIleAspIleAlaArgGluAsnAsnProAspTrp	100
Db	376	TCCCAACGCAACTCATCTTAAAGATATCCATATCGCGCGCAAAACAAACCCGATTTG	435
QY	101	IleLeuSerAsnLysGlnValGlyValCysTyrValAspLeuPheAlaGlyAspLeu	120
Db	436	ATTITGTCCAAACAACAAGTCGCGCGCTGTGCTACGTTGATTGTTGTTGCGGGGATTG	495
QY	121	LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrIleHisLeu	140
Db	496	AAGGCTTTGAAGATAAAATTCCTTATTTCAAGAGCTTGGTTGACTATCTGCACCTG	555
QY	141	MetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyr	160
Db	556	ATGCGCGTGTAAATGCCCTTGAAGCAACGACGCGCGCTATGCGTCAGCAGCTAC	615
QY	161	ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAlaLeu	180
Db	616	CGCATGTCTCAATCCGGCACTGGGCACAATAGCGACTTTCGCGAAGTCAATTCGCGCTG	675
QY	181	HisGluAlaGlyIleSerAlaValAlaValAspPheIlePheAsnHisThrSerAsnGluHis	200
Db	676	CACGAAGCGCGCATTTCCGCGCTCGTCGATTTTATCTTTCAACCAACACCTCCACGACAC	735
QY	201	GluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePhe	220
Db	736	GAATGGCGCAACGCTGGCGCGCGCGCGCGCTTTTCGACAATTTCTACTATATTTC	795
QY	221	ProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGln	240
Db	796	CCCAGCGCGCGATGCCGACCAATACGACGCGACCCCTTCGCGAAATCTTCCCGACACAG	855
QY	241	HisProGlyGlyPheSerGlnLeuGluAspGlyArgTrpValTrpThrPheAsnSer	260
Db	856	CACCGCGCGCTTCTCGCACTCGAAGACGACGCTGGTGTGGACGCACTTCAATTC	915
QY	261	PheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheArgAlaMetAlaGlyGluMet	280
Db	916	TTCCAATGGGACTTGAATTAACGCAACCCCGTGGGTATTTCGCGCAATCGCGCGGCAATG	975
QY	281	LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTrp	300
Db	976	CTGTCTCTTGCACACTGGCGGTGTGACATCTCTGATGATGATGCGGTTCCTTTATTGG	1035
QY	301	LysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPhe	320
Db	1036	AAACAATGGGCAACAGCTGCAAAACCTGCGCGAGCGCACGCCCTCATCCGCGCGTTC	1095

ORGANISM	synthetic construct	artificial sequences.
REFERENCE	1	
AUTHORS	Bengs H., Polakowski T., Heid, A. and Gallert, K.C.	
TITLE	Novel immobilizable amylase, the use thereof and method for producing poly(1,4-alpha-glucan)	
JOURNAL	Patent: WO 0125449-A 6 12-APR-2001;	
AXIVA GmbH (DE)		
FEATURES	Location/Qualifiers	
source	1..6878	
	/organism="synthetic construct"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:32630"	
	/note="Beschreibung der künstlichen Sequenz: Expressionsvektor pGEX-4T-1-Amsu"	
BASE COUNT	1673 a 1800 c 1766 g 1639 t	
ORIGIN		
Alignment Scores:		
Pred. No.:	3,36e-286	Length: 6878
Score:	3408.00	Matches: 636
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	100.00%	Indels: 0
DB:	6	Gaps: 0
US-09-843-007A-2 (1-636) x AX107015 (1-6878)		
QY	1 MetLeuThrProThrGlnGlnValGlyLeuLeuGlnTyrLeuLysThrArgIleLeu 20	
DB	945 ATGTTGACCCACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1004	
QY	21 AspIleTyrThrProGluGlnArgAlaGlyIleGluLysSerGluAspTrpArgGlnPhe 40	
DB	1005 GACATCTACACGCGCGAACACGACGCGCGGACATCGAAAAATCCGAAGACTGCGCGCAGTTT 1064	
QY	41 SerArgGlnMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGly 60	
DB	1065 TCGCGCGCGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1124	
QY	61 AsnAsnGluAlaLeuLeuProMetLeuGluMetLeuLeuAlaGlnAlaTrpGlnSerTyr 80	
DB	1125 AACACGAAAGCCCTGCTGCTATGCTGMAATGCTGCTGCGCGACGACGACGACGACG 1184	
QY	81 SerGlnArgAsnSerSerLeuLysAspIleAspIleAlaArgGluAsnAsnProAspTyr 100	
DB	1185 TCCACGCGCACTCTCTCTTAAAGATATGATATGCGCGCGCGCGCGCGCGCGCGCG 1244	
QY	101 IleLeuSerAsnLysGlnValGlyValCysTyrValAspLeuPheAlaGlyAspLeu 120	
DB	1245 ATTTTGTCCAAACAAACAGTTCGGCGCGTGTGCTACGTGCTGCTGCTGCTGCTGCTG 1304	
QY	121 LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeu 140	
DB	1305 AAGGGCTTCAAGAGATAAATTCCTTATTTTCAAGAGCTTGGTTTGTGACTATATCTGAC 1364	
QY	141 MetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyr 160	
DB	1365 ATGCCGCTGTTTAAATGTCCTGAGGCGAAAGCGACGCGCGCTATGCGGTGACGAGCTAC 1424	
QY	161 ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaLeu 180	
DB	1425 CGCGATGTCAATCCGCGCATGCGCGCAATAGCGCGCTTTCGCGCGAGTCAATGCTGCGCGCT 1484	
QY	181 HisGluAlaGlyLysSerAlaValValAspPheIlePheAsnHisThrSerAsnGluHis 200	
DB	1485 CAGAGAGCGCGCATTTCCGCGCTGCTGCGCTGCTGCGCTGCTGCGCTGCTGCGCTG 1544	
QY	201 GluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePhe 220	
DB	1545 GAATGGCGCGCAACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1604	
QY	221 ProAspArgArgMetProAspGlnTyrAspArgThrLeuAspGluIlePheProAspGln 240	
QY	321 AsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHis 340	
DB	1096 AATGCGGTATGCGGTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1155	
QY	341 ProAspGlnValValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnProLeu 360	
DB	1156 CCGGACCAAGTCGTCCTATACATCGGCGAGGACGAAATGCCAAATCGGTTTCAACCCCGCTG 1215	
QY	361 GlnMetAlaLeuLeuTyrAsnThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAla 380	
DB	1216 CAAATGCGATTTGTTGGACACCTTGGCAGCGCGGAGTCAACTGCTCCATCAGCGG 1275	
QY	381 LeuThrTyrArgHisAsnLeuProGluHisThrAlaTrpValAsnTyrValArgSerHis 400	
DB	1276 CTGACCTACCGGCACAACTGCGCGCGGATACCGCTGCGTCAACTACGTCGCGAGCCAC 1335	
QY	401 AspAspIleGlyThrThrPheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyr 420	
DB	1336 GAGCATCGCTGGACGTTTCCGATGAGACGCGCGGATATCTGGGCAATAGCGCGCTAC 1395	
QY	421 AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg 440	
DB	1396 GACCAACCGCAATTCCTCAACCGCTTCTGCTCAACCGCTTTCGACGCGCGCTTCGCTCGT 1455	
QY	441 GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla 460	
DB	1456 GCGGTACCGTTCATACCAACCCAGCAGCGCGGACTGCCGCTGCTGCTGCTGCTGCTGCT 1515	
QY	461 AlaLeuValGlyLeuAlaGlnAspAspProHisAlaValAspArgIleLysLeuLeuTyr 480	
DB	1516 GCATTGCTGGCTTGGCGCAAGACGATCCCAACGCGCTTGACCGCATCAAACTCTTGATC 1575	
QY	481 SerIleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr 500	
DB	1576 AGCATTTGCTTTAGTACCGCGCGCTGCGCGCTGATTTACCTAGGCGACGAGTGGGTACG 1635	
QY	501 LeuAsnAspAspTyrSerGlnAspSerAsnLysSerAspAspSerArgTyrAlaHis 520	
DB	1636 CTCATGACGACGACGCTGCTGCAAGACAGACCAATAGACGACGACGACGCGTGGCGGCAC 1695	
QY	521 ArgProArgTyrAsnGluAlaLeuTyrAlaGlnAsnAspProSerThrAlaAlaGly 540	
DB	1696 CGTCCGCGCTACCAACGAAGCCCTGTACGCGCAACGCAAGTCCGTCGACCGCGCGCGG 1755	
QY	541 GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp 560	
DB	1756 CAATCTATCAGGCTTGCCTCATATGATGCGTCCGCGCAAGCAATCCGCGCTTCGAC 1815	
QY	561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleGlyTyrIleArgAsn 580	
DB	1816 GCGCGCGAGCTGGTTACATTCACACCAACCAACGACACATCATCGCTGCTACATCGCGAAC 1875	
QY	581 AsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThr 600	
DB	1876 AATGCGCTTTGGCATTCGGTAACTTCGCGCAATATCCGCAACCGTTACCGCGCATACC 1935	
QY	601 LeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyLysThrValSerLeuAsn 620	
DB	1936 CTGCAAGCCATCGCTTCAAGGCGCACGACCTCATCGCTGGCGAAATCTGTACGCTGAAT 1995	
QY	621 GlnAspLeuThrLeuGlnProTyrGlnValMetTrpLeuGluIleAla 636	

Db 1605 CCGACCGCGGATGCCGACCAATACGACGACCTCTGGCGAATCTTCCCGGACCG 1664  
Qy 241 HisProGlyGlyPheSerGlnLeuGluAspGlyArgTrpValTrpThrPheAsnSer 260  
Db 1665 CACCGCGCGGCTTCTCGCAACTGGAGACGCGCGTGGGTGGTGGACGACCTTCAATTC 1724  
Qy 261 PheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheArgAlaMetAlaGlyGluMet 280  
Db 1725 TTCATATGGGACCTTGAATATACGCAACCGGTGGGTATTCGCGCAATGGGGGGAATG 1784  
Qy 281 LeuPheLeuAlaAsnLeuGlyValAspLeuArgMetAspAlaValAlaPheIleTrp 300  
Db 1785 CTGTTCCTTCCCACTTTGGCGTTTGACATCTCTGCTATGATGCGGTGGCTTATTTGG 1844  
Qy 301 LysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuAlaPhe 320  
Db 1845 AACCAATGGGACAACTGGCAAACTCTGGCGGCGGCGACCTCATCCGGCGGCTTC 1904  
Qy 321 AsnAlaValMetArgIleAlaProAlaValPhePheLysSerGluAlaIleValHis 340  
Db 1905 ATGCGGTTATGCGTATTCGCGCGCGCGCTGTCTTCAATCCGAAGCCATCGTCCAC 1964  
Qy 341 ProAspGlnValValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnProLeu 360  
Db 1965 CCCGACCAAGTCGTCCAATACATCGGCGGACGAATGCCAAATCGGTTCACACCCCTG 2024  
Qy 361 GlnMetAlaLeuLeuTyrAsnThrLeuAlaThrArgGluValAsnLeuHisGlnAla 380  
Db 2025 CAATGGCATTTGTTGGACACCTTTGCGACGCGGCGAGTCAACCTGCTCCATCAGCG 2084  
Qy 381 LeuThrTyrArgHisAsnLeuProGluHisThrAlaTrpValAsnTyrValArgSerHis 400  
Db 2085 CTGACCTACCGCCACCACTGCCCGGACGATACCGCTGGGTCAACTAGTCCGCGGCGAC 2144  
Qy 401 AspAspIleGlyTrpThrPheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyr 420  
Db 2145 GACGACATCGCTGGACGCTTTGCGGATGAACGCGGCGATATCTGGGCTAAGCGGTAC 2204  
Qy 421 AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg 440  
Db 2205 GACCAACCGCAATCTCTCAACGCTTCTTCGTCAACGCTTTGCGCGGCGAGCTTCGCT 2264  
Qy 441 GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla 460  
Db 2265 GCGGTACCGCTTCCAATACACCCAGACAGCGGCGACTGCCGTGTGAGTGGTACAGCGCG 2324  
Qy 461 AlaLeuValGlyLeuAlaGlnAspProHisAlaValAspArgIleLysLeuLeuTyr 480  
Db 2325 GCATTTGGTGGCTTGGCGCAAGACGATCCCAACGCGCTTGACCGCATCAACTCTTGATC 2384  
Qy 481 SerIleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr 500  
Db 2385 AGCATTTGCTTTGAGTACCGCGGCTCTGCGCGCTGATTTACCTAGGCGACGAAAGTGGTACG 2444  
Qy 501 LeuAsnAspAspAspTrpSerGlnAspSerLysSerAspAspSerArgTrpAlaHis 520  
Db 2445 CTCATATGACGAGCTGTGGTGGCAAGACGATCCCAACGCGCTTGACCGCATCAACTCTTGATC 2504  
Qy 521 ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGly 540  
Db 2505 CGTCCGCGCTACACGAGCCCTGTACGCGCAACGCAACGATCCGCTCGACGCGCGCGG 2564  
Qy 541 GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp 560  
Db 2565 CAATCTATCAGGCTTGGCGCATATGATTTGCGTCCGCGCAAGCAATCCGCGCTTCGAC 2624  
Qy 561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArgAsn 580  
Db 2625 GCGCGGAGGCTGGTTACATTCACACCAACACAGCACATCATCGCTACATCCGCAAC 2684  
Qy 581 AsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThr 600

Db 2685 AATGGCGCTTTGGCATTCGGTAACCTTCAGCGAATATCCGCAACCGTTACCGCGCATACC 2744  
Qy 601 LeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyLysThrValSerLeuAsn 620  
Db 2745 CTGCAAGCCATGCCCTTCAGGCGGACAGACCTCATCGTGGCAAACTGTCACGCTGAT 2804  
Qy 621 GlnAspLeuThrLeuGlnProTyrGlnValMetTrpLeuGluIleAla 636  
Db 2805 CAGGATTTGACGCTTCAGCCCTATCAGGTCAATGCTGCTCGAATCGCC 2852  
RESULT 4  
AX107014 1910 bp DNA linear PAT 30-APR-2001  
LOCUS Sequence 5 from Patent WO0125449.  
DEFINITION AX107014  
ACCESSION AX107014.1 GI:13922548  
VERSION  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM artificial sequences.  
REFERENCE 1  
AUTHORS Bengs,H., Polakowski,T., Held,A. and Gallert,K.C.  
TITLE Novel immobilizable amylosucrase, the use thereof and method for  
JOURNAL producing poly(1,4-alpha-glucan)  
Patent: WO 0125449-A 5 12-APR-2001;  
Axiva GmbH (DE)  
FEATURES Location/Qualifiers  
source 1..1910  
/organism="synthetic construct"  
/mol type="genomic DNA"  
/db xref="taxon:32630"  
/note="PCR-Amplikon, kodierend fur verkurzte amylosucrase (AMSuS)"  
BASE COUNT 447 a 595 c 474 g 387 t 7 others  
ORIGIN  
Alignment Scores:  
Pred. No.: 8,36e-283 Length: 1910  
Score: 330.00 Matches: 626  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.84% Mismatches: 0  
Query Match: 98.59% Indels: 0  
DB: Gaps: 0  
US-09-843-007A-2 (1-636) x AX107014 (1-1910)  
Qy 10 LeuIleLeuGlnTyrLeuLysThrArgIleLeuAspIleTyrThrProGluGlnArgAla 29  
Db 10 ATGATTTTACAGTACCTCAAAACACGCGATCTTGACATCTACACGCCCGGACAGCGCGCC 59  
Qy 30 GlyIleGluLysSerGluAspTrpArgGlnPheSerArgArgMetAspThrHisPhePro 49  
Db 70 GGCATCGAATAATCCGAGACTGGCGGCACTTTTGGCGCGCATGGATACGCAATTCGCC 129  
Qy 50 LysLeuMetAsnGluLeuAspSerValTyrGlyAsnAsnGluAlaLeuLeuProMetLeu 69  
Db 130 AAACGTGATGAACCAACTCGACGCGTGTACGCGCAACCAACGAGCCCTGCTGCCCTATGCTG 189  
Qy 70 GluMetLeuLeuAlaGlnAlaTrpGlnSerTyrSerGlnArgAsnSerSerLeuLysAsp 89  
Db 190 GAATATGCTGTGGCGCGAGGCAATGGCAAGACTATTCCTCAACGCACTCATCTTAAAGAT 249  
Qy 90 IleAspIleAlaArgGluLeuAsnAsnProAspTrpIleLeuSerAsnLysGlnValGlyGly 109  
Db 250 ATCGATATCGCGCGAAGAAACAAACCCGATTTGGATTTTGTCCAAACAAACAGTCGCGCGC 309  
Qy 110 ValCysTyrValAspLeuPheAlaGlyAspLeuLysGlyLeuLysAspLysIleProTyr 129  
Db 310 GTGTGCTACGTTGATTTGTTTGGCGGCGATTTGAAGGCTTTGAAGATATAAATTCCTTAT 369  
Qy 130 PheGlnGluLeuGlyLeuThrTyrLeuHisLeuMetProLeuPheLysCysProGluGly 149  
Db 370 TTTCAAGAGCTTGGTTTGACTTAICTGCACCTGATCGCGTGTTTAAATGCTCCCTGAGGC 429

150 LysSerAspGlyGlyTyrAlaValSerSerTyrArgAspValAsnProAlaLeuGlyThr 169  
 170 IleGlyAspLeuArgGluValIleAlaLeuHisGluAlaGlyIleSerAlaValVal 189  
 190 AspPheIlePheAsnHisThrSerAsnGluHisGluTrpAlaGlnArgCysAlaAlaGly 209  
 210 AspProLeuPheAspAsnPheTyrIlePheProAspArgMetProAspGlnTyr 229  
 230 AspArgThrLeuArgGlnIlePheProAspGlnHisProGlyGlyPheSerGlnLeuGlu 249  
 250 AspGlyArgTrpValTrpThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsn 269  
 270 ProTrpValPheArgAlaMetAlaGlyGlyMetLeuPheLeuAlaAsnLeuGlyValAsp 289  
 290 IleLeuArgMetAspAlaValAlaPheIleTrpLysGlnMetGlyThrSerCysGlyAsn 309  
 310 LeuProGlnAlaHisAlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaPro 329  
 330 AlaValPhePheIleSerGluAlaIleValHisProAspGlnValValGlnTyrIleGly 349  
 350 GlnAspGlyCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTrpAsnThrLeu 369  
 370 AlaThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGlu 389  
 390 HisThrAlaTrpValAsnTyrValArgSerHisAspAlleGlyTrpThrPheAlaAsp 409  
 410 GluAspAlaIleValLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsnArgPhe 429  
 430 PheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsnProSer 449  
 450 ThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAlaGlnAspAsp 469  
 470 ProHisAlaValAspArgIleLysLeuLeuTyrSerIleAlaLeuSerThrGlyGlyLeu 489  
 490 ProLeuIleTyrLeuGlyAspGluValGlyThrLeuAsnAspAspAspTrpSerGlnAsp 509  
 510 SerAsnLysSerAspAspSerArgTrpAlaHisArgProArgTyrAsnGluAlaLeuTyr 529

1510 AGCAATAAGACGACGACGACGCGTGGCGCGACCGCTCGCGCTACCAAGACCGCTGTAC 1569  
 1530 AlaGlnArgAsnAspProSerThrAlaAlaGlyGlnIleTyrGlnGlyLeuArgHisMet 1549  
 1570 GCGCAACCAACGATCCGTCACCGACGCGCGGCAATCTATCAGGGCTTCGGCCATTG 1629  
 1590 IleAlaValArgGlnSerAsnProArgPheAspGlyGlyArgLeuValThrPheAsnThr 1599  
 1630 ATTGCGCTCCGCCAAGCAATCCGCTTCGACCGCGCGAGCTGGTTACATTCAACACC 1689  
 1650 AsnAsnLysHisIleIleGlyTyrIleArgAsnAsnAlaLeuAlaPheGlyAsnPhe 1589  
 1690 AACACAAAGCACATCATCGCTACATCGCAACAAATGGCGCTTTGGCAATTCGTAACTTC 1749  
 1710 SerGluTyrProGlnThrValThrAlaHisThrLeuGlnAlaMetProPheLysAlaHis 1609  
 1750 AGCGAATATCCGCAACCGTTACCGCGCATACCTCGCAAGCCATCGCTTCAAGGCGCAC 1809  
 1770 AspLeuIleGlyGlyIleValSerLeuAsnGlnAspLeuThrLeuGlnProTyrGln 1629  
 1810 GACCTCATCGGTGGCAAACTGTACGCTGAATCAGGATTTGACGCTTCAGCCCTATCAG 1869  
 1830 ValMetTrpLeuGluIleAla 636  
 1870 GTCAATGGCTCGAAATCGCC 1890

RESULT 5  
 LOCUS AX107016 6851 bp DNA linear PAT 30-APR-2001  
 DEFINITION Sequence 7 from Patent WO0125449.  
 ACCESSION AX107016  
 VERSION AX107016.1 GI:13922550  
 KEYWORDS synthetic construct  
 SOURCE synthetic construct  
 ORGANISM artificial sequences.  
 REFERENCE 1  
 AUTHORS Bengs, H., Polakowski, T., Held, A. and Gallert, K.C.  
 TITLE Novel, immobilizable amylase, the use thereof and method for  
 JOURNAL producing poly(1,4-alpha-glucan)  
 Patents: WO 0125449-A 7 12-APR-2001;  
 Axiva GmbH (DE)  
 FEATURES  
 Location/Qualifiers  
 1..6851  
 /organism="synthetic construct"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:32630"  
 /note="Beschreibung der kunstlichen  
 Sequenz: Expressionsvektor pGEX-4T-1-Amsus"  
 BASE COUNT 1668 a 1791 c 1759 g 1633 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 5,04e-282 Length: 6851  
 Score: 3360.00 Matches: 626  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 99.84% Mismatches: 0  
 Query Match: 98.59% Indels: 0  
 Gaps: 6  
 DB:

US-09-843-007a-2 (1-636) x AX107016 (1-6851)

10 LeuIleLeuGlnTyrIleLysThrArgIleLeuAspIleTyrThrProGlnArgAla 29  
 945 ATGATTTTACAGTACCTCAAAACACGCACTCTTGGACATCTACACGCCCGCAACGCGCGC 1004  
 30 GlyTleGluLysSerGluAspTrpArgGlnPheSerArgMetAspThrHisPhePro 49  
 1005 GGCATCGAAATCCGAAGACTGGCGGCGAGTTTTCGCCCGCATGGATACGATTCTCCC 1064  
 50 LysLeuMetAsnGluLeuAspSerValTyrGlyAsnAsnGluAlaLeuProMetLeu 69

Db 1065 AAACTGATGAACGAACCTGACAGCGTGTACGGCAACAACGAAGCCCTGCTGCTATGCTG 1124  
Qy 70 GluMetLeuAlaGlnAlaTirpGlnSerTyrSerGlnArgAsnSerSerLeuIysAsp 89  
Db 1125 GAAATGCTGCTGGCGAGCATGGCAAGCATTTCCCAACGCAACTCATCTCTTAAAGAT 1184  
Qy 90 IleAspIleAlaArgGluAsnProAspTirpIleLeuSerAsnLysGlnValGly 109  
Db 1185 ATCGATATGCGCGCGAACAACCCGATTTGGATTTTGTCCAAACAAACATCGCGCGC 1244  
Qy 110 ValCysTyrValAspLeuPheAlaGlyAspLeuLysGlyLeuLysAspLysIleProTyr 129  
Db 1245 GTGGCTACCTGATTTGTTTCCCGCGCATTTGAAGCGCTTGAAGATTAATTCCTTAT 1304  
Qy 130 PheGlnGluLeuGlyLeuThrTyrLeuHisLeuMetProLeuPheLysCysProGluGly 149  
Db 1305 TTTCAGAGCTTGGTGTGATCTTATCGCACTGATGCGCTGTTTAAATGCGCTGAAGC 1364  
Qy 150 LysSerAspGlyGlyTyrAlaValSerSerTyrArgAspValAsnProAlaLeuGlyThr 169  
Db 1365 AAAAGCGAGCGCGGTATGCGTACAGCTACCGCATTCATCCGCGCATGGGCA 1424  
Qy 170 IleGlyAspLeuArgGluValIleAlaLeuHisGluAlaGlyIleSerAlaVal 189  
Db 1425 ATAGCGGACTTGGCGGAGTCTGCGCTGCGACGAAGCGCGCATTTCCGCGCTGTC 1484  
Qy 190 AspPheIlePheAsnHisThrSerAsnGluHisGluTirpAlaGlnArgCysAlaGly 209  
Db 1485 GATTTTATCTTCAACCAACCTTCCACGAACAGAAATGGCGCAACGCTGCGCGCGC 1544  
Qy 210 AspProLeuPheAspAsnPheTyrTyrIlePheProAspArgArgMetProAspGlnTyr 229  
Db 1545 GACCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCCCAATAC 1604  
Qy 230 AspArgThrLeuArgGluIlePheProAspGlnHisProGlyGlyPheSerGlnLeuGlu 249  
Db 1605 GACCGCACCTCGCGAATCTTCCCGACGACGACCGCGCGCTTCTCGCAACTGGAA 1664  
Qy 250 AspGlyArgTirpValTirpThrPheAsnSerPheGlnTirpAspLeuAsnTyrSerAsn 269  
Db 1665 GACGGACGCTGGGTGGAGCACTTCATTTCTTCATGCGACTTGAATACAGCAAC 1724  
Qy 270 ProTirpValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAsp 289  
Db 1725 CCGTGGGTATTCGCGCAATGGCGGCAATGCTGTTCTTCCCAACTTGGCGCTGAC 1784  
Qy 290 IleLeuArgMetAspAlaValAlaPheIleTirpLysGlnMetGlyThrSerCysGluAsn 309  
Db 1785 ATCTCGGTATGATCGGTGCGTTCCTTTATTTGGAACAATGGGCAAGCTGCGAACA 1844  
Qy 310 LeuProGlnAlaHisAlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaPro 329  
Db 1845 CTGCGCAGCGCGCACGCCCTCATCGCGCGTTCATGCGTTCATGCGTATTCGCGCGC 1904  
Qy 330 AlaValPhePheLysSerGluAlaIleValHisProAspGlnValValGlnTyrIleGly 349  
Db 1905 GCGGTGTTCTTCAAAATCCGAAGCATCGTCCACCGCGACCAAGTCTCCAAATACATCGG 1964  
Qy 350 GlnAspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTirpAsnThrLeu 369  
Db 1965 CAGGAGCAATGCCAATCGGTTCACACCCCTCGCAATGCGATTTGTGGACACCTT 2024  
Qy 370 AlaThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGlu 389  
Db 2025 GCCACGCGGAGTCAACCTGCTCCATCAGCGCTGACCTACCGCACCAACCTGCCCGAG 2084  
Qy 390 HisThrAlaTirpValAsnTyrValArgSerHisAspAspIleGlyTirpThrPheAlaAsp 409  
Db 2085 CATACCGCTGGTCAACTACGTCCGACGACGACATCGCTGGACGTTTGGCAT 2144  
Qy 410 GluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsnArgPhe 429

Db 2145 GAAGACCGCGCATATCTGGGCATTAAGCGGTACGACACCGCCAAATCTCTCAACGCTTC 2204  
Qy 430 PheValAsnArgPheAspGlySerPheAlaArgGlyValPropheGlnTyrAsnProSer 449  
Db 2205 TTCTGCAACCGTTTCGACGCGCAGCTTCGCTCGTGGCTACCGTTCCAATCAACCCAGC 2264  
Qy 450 ThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAlaGlnAspAsp 469  
Db 2265 ACAGCGACCTCGGTGTCAGTGGTACAGCGCGGCATTTGGCTGGCCAGACCAT 2324  
Qy 470 ProHisAlaValAspArgIleLysLeuLeuTyrSerIleAlaLeuSerThrGlyGlyLeu 489  
Db 2325 CCCACGCGGTGACCGCATCAAACTCTTGTACAGCATGCTTTGAGTACCGCGGCTG 2384  
Qy 490 ProLeuIleTyrLeuGlyAspGluValGlyThrLeuAsnAspAspAspTirpSerGlnAsp 509  
Db 2385 CCGTGTATTTACCTAGCGAGCAAGTGGGTACGCTCAATACGACGACTGGTTCGCAAGAC 2444  
Qy 510 SerAsnLysSerAspAspSerArgTirpAlaHisArgProArgTyrAsnGluAlaLeuTyr 529  
Db 2445 AGCATAAGACGACGACGACGCGTGGCGGACCGCTGCGCGCTACACGAAGCCCTGAT 2504  
Qy 530 AlaGlnArgAsnAspProSerThrAlaAlaGlyGlnIleTyrGlnGlyLeuArgHisMet 549  
Db 2505 CGCAACGCAACGATTCGCTCGACCGCGCGCGGCAATCTATCAGGCTTTCGCGCATATG 2564  
Qy 550 IleAlaValArgGlnSerAsnProArgPheAspGlyGlyArgLeuValThrPheAsnThr 569  
Db 2565 ATTCCGCTCCGCCAAGCAATCCGCGCTTCGACGCGCGGCTTGGCATTCGTAACCTC 2684  
Qy 570 AsnAsnLysHisIleIleGlyTyrIleArgAsnAsnAlaLeuLeuAlaPheGlyAsnPhe 589  
Db 2625 AACCAACGACACATCATCGGTACATCCGCAACATCGCTTTCGCGCTTTCGTAACCTC 2684  
Qy 590 SerGluTyrProGlnThrValThrAlaHisThrLeuGlnAlaMetPropheLysAlaHis 609  
Db 2685 AGCGAATATCCGCAACCGTTACCGCGCTACCTTCGACGCGCTTTCGCGCGCTTCAAGCGG 2744  
Qy 610 AspLeuIleGlyGlyLysThrValSerLeuAsnGlnAspLeuThrLeuGlnProTyrGln 629  
Db 2745 GACCTCATCGTGGCAAACTGTGAGCTGATCAGGATTTGAGCTTTCGCGCTTTCGCGCTATCAG 2804  
Qy 630 ValMetTirpLeuGluIleAla 636  
Db 2805 GTCATGTGGCTCGAAATCGCC 2825

RESULT 6  
AY099334 1967 bp DNA linear BCT 13-JAN-2003  
LOCUS Neisseria meningitidis strain 93246 amylosucrase Ams (ams) gene,  
DEFINITION complete cds.  
ACCESSION AY099334  
VERSION AY099334.1 GI:27728141  
KEYWORDS  
SOURCE  
ORGANISM  
Neisseria meningitidis  
Neisseria meningitidis  
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
Neisseriaceae; Neisseria.  
REFERENCE  
AUTHORS  
TITLE  
Zhu, P., Teang, R.-S.W. and Tsai, C.M.  
Nonencapsulated Neisseria meningitidis Strain Produces Amylopectin  
from Sucrose: Altering the Concept for Differentiation between N.  
meningitidis and N. polysacchara  
J. Clin. Microbiol. 41 (1), 273-278 (2003)  
JOURNAL  
PUBMED 12517860  
REFERENCE 2 (bases 1 to 1967)  
AUTHORS Zhu, P.  
TITLE Direct Submission  
JOURNAL Submitted (23-APR-2002) LBP/DBPAP, CBER, 8800 Rockville Pike,  
Bethesda, MD 20892, USA  
FEATURES  
Location/Qualifiers  
1..1967  
/organism="Neisseria meningitidis"

```

/mol_type="genomic DNA"
/strain="93246"
/db_xref="taxon:487"
21..1931
/gene="ams"
21..1931
/gene="ams"
/codon_start=1
/transal_table=11
/product="amylase Ams"
/protein_id="AA51152.1"
/db_xref="GI:27728142"
/translation="MELPTQQVGLIILQYLKTRILDIYTPORAGIEKSEDWQFQSRM
DTHFPLKMLLEDSVYGNALPLMLLELAQWQSYQSASLKNIDIERENNPWIL
SNQVGGCYVDLFDGLKDKHIFQELGLTYLHLMPLFKPGKSDGDAYVSSY
RDVNPALGIDRLREIVIAALHEAGISAVVDIFPNHSEHEWAQRCAAGDPLDFNY
IPDRMPQIDRLREIPDHPGPGFSQLEDRQWWTTFNSFOWDLNYSNFWVRAM
AGSMFLANGLDRLMDAVAFIWKOMGTSCENLPOAHALIRAFNAWMSIAAPAVFFK
SEALVHPDQVQYIGDEQCIQVNPLOMALIWNLTATREVNLHQAITYRHLPSHTA
WNYVPSHDIQWTPADEDAAYLIGSYDHRQFLARFFVNRFDGSPARGVPQYQNP
GCKRSGTAAAGLAQNDPFAVSRILKLIYSIALSTGGPLIYIGDEVGTFPNDGQVQ
DSNKSDDSRWAHPRNEALYACRNDPSTAAGQIYQGLRMIYAVQSNRPFQDGLVT
FNTNNKHITGYRNALLAFGNFSEYPTQVTAHTLOAMPFKADHLIGKTVSLNODLT
LOFYQVWLEIA"
BASE COUNT 463 a 626 c 490 g 388 t
ORIGIN

Alignment Scores:
Pred. No.: 7,17e-281 Length: 1967
Score: 3338.00 Matches: 622
Percent Similarity: 98.90% Conservative: 7
Best Local Similarity: 97.80% Mismatches: 7
Query Match: 97.95% Indels: 0
DB: 1 Gaps: 0

US-09-843-007A-2 (1-636) x AY099334 (1-1967)

QY 1 MetLeuThrProThrGlnValGlyLeuIleLeuGlnTyrLeuLysThrArgIleLeu 20
DB 21 ATGTTGACCCCAACGAGCAAGTCGGTGTGATTTTACAGTACCTCAAAACACGCACTTG 80
QY 21 AspIleTyrThrProGlnValGlyLeuIleGlyLeuSerGlnAspTyrArgGlnPhe 40
DB 81 GACATCTACACGCCCGCAACAGCGCGCGCATCGAATAATCGAAGACTGGCGCGAGTTT 140
QY 41 SerArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGly 60
DB 141 TCGCGCGCGCATGATACGCAATTTCCCAAACTGATGAACTCGACAGCGGTACCGGC 200
QY 61 AsnAsnGluAlaLeuLeuProMetLeuGluMetLeuAlaGlnAlaTyrGlnSerTyr 80
DB 201 AACACAGAGACCCCTGCTGCTATGCTGGAATGCTGCGCGCAGGCATGGCAGACTAT 260
QY 81 SerGlnArgAsnSerLeuLysAspIleAspIleAlaArgGluAsnAsnProAspTyr 100
DB 261 TCCCAACGAGCGCGCTCCCTGGAACATGATGATCAACGCGAACAACCCCGATTGG 320
QY 101 IleLeuSerAsnLysGlnValGlyGlyValCysTyrValAspLeuPheAlaGlyAspLeu 120
DB 321 ATTATTGTCACAAACAAAGTTGGCGGTGTGTCTGATGTTGTTGTCGCGCGCATTTA 380
QY 121 LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrIleuHisLeu 140
DB 381 AAAGGCTGAAAGACAAATTCATTATTTTCAAGAGCTTGGTTGACTTATCTGCACCTG 440
QY 141 MetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAlaValSerSerTyr 160
DB 441 ATGCGCGCTGTTAAATAGCCCTGAAGCAAAAGCGACGCGCGCTATCGCGTACGAGCTAC 500
QY 161 ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaLeu 180
DB 501 CCGCATGTCAATCCGCACTGGGCACAAATAGGCGACTTGGCGAAGTCAATCTGCGCTG 560

101 HisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSerAsnGluHis 200
561 CACGAAGCGCGCATTTCCGCGCTGCGATTTTATCTTCAACACACACCTCCCAACGACAC 620
201 GluTyrAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePhe 220
621 GAATGGCGCAACGCTGCGCGCGCGCGCGCGCTTTTCGACAAATTTCTACTATATTTTC 680
221 ProAspArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGln 240
681 CCGCAGCGCGGATGCCCGACCAATACGACCGCACCTCTGCGGAAATCTTCTCCCGACAC 740
241 HisProGlyGlyPheSerGlnLeuGluAspGlyArgTyrValTyrThrThrPheAsnSer 260
741 CACCGCGCGCGCTTCTGCAACTGGAAGACGCGCGCTGCTGCGGACGCTTCATATTC 800
261 PheGlnTyrAspLeuAsnTyrSerAsnProTyrValPheAspIleAlaMetAlaGlyLeuMet 280
801 TTCCAAATGGGACTTGAATTAACAGCAACCGCTGGGTATTTCCGCGCAATGGCGCGGAAATG 860
281 LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTyr 300
861 ATGTTCTCTGCAACTTGGCGGTTCACATCTTGGCTATGATGATGCGGTTCGCTTTATTGG 920
301 LysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPhe 320
921 AAACAATGGGACCAAGCTGGCAAACTGCGCAGGCGCGACGCCCTCATCGCGCGTTC 980
321 AsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHis 340
981 AATGCGGTATGCGTATCGCGCGCGCGCGCGCTGTTCTTCAAAATCCGAAGACCATCGTCCAC 1040
341 ProAspGlnValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnProLeu 360
1041 CCGACCAAGTCGTCCAATATATCGGCGAGCAGCAATGCCAAATCGCTACAAACCCCTG 1100
361 GlnMetAlaLeuLeuTyrAsnThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAla 380
1101 CAATGGCATTTGTTGGAAACACCCCTTCCCAACGCGCGAAGTCAACCTGCTCCATCAGCGC 1160
381 LeuThrTyrArgHisAsnLeuProGluHisThrAlaTyrValAsnTyrValArgSerHis 400
1161 CTGACTTACCGCCCAACCTGCGCGAGCATATCTGCTGGGTCAACTAGCTCGCGAGCCAC 1220
401 AspAspIleGlyTyrThrPheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyr 420
1221 GACGACATCGGCTGACGCTTTGCGGATGAGACGCGGCATATCTGGGCATTAAGCGGTAC 1280
421 AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg 440
1281 GACCACCGCCAGTTCTTCAACCGCTTCTTCTGCAACCGTTTCGACGCGCAGCTTCGCGCGC 1340
441 GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla 460
1341 GCGGTACCTTCCAAATACACCCCAACACGCGCGACCTGCGCGCTACGCGGTACGTCGCC 1400
461 AlaLeuValGlyLeuAlaGlnAspProHisAlaValAspArgIleLysLeuLeuTyr 480
1401 GCATTGGCAGGCTTGGCTCAAAATGACCCCGCACCGCTCAGCGCATCAAACTTTTGTAC 1460
481 SerIleAlaLeuSerThrGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr 500
1461 AGCATCGCATTTGAGTACCGCGCGGCTGCCCTGATTTTACCTTCGCGCAGCAAGTGGCAGC 1520
501 LeuAsnAspAspAspTyrSerGlnAspSerAsnLysSerAspAspSerArgTyrAlaHis 520
1521 CCCAACGACGCGGTGGGTGCAAGCAGCAACAAAGACGACGACGACGCGCTTGGCGCGAC 1580
521 ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGly 540
1581 GTCGCGCGCTATACGAAGCCCTGTACGCGCAACGACGATTCGCTCGCGCGCGCGCGG 1640
541 GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp 560

```



```

Db      1641 CAAATCTATCAGGCTTGGCCATATGATTCGCTCGCAAGCAATCGCGCTTCGAC 1700
Qy      561 GlyGlyArgLeuValThrPheSerThrAsnAsnLysHisIleGlyTyrIleArgAsn 580
Db      1701 GCGGCGAGCTGGTACCTTCAACCAACCAACCAACCAACCAACCAACCAACCAAC 1760
Qy      581 AsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThr 600
Db      1761 AACGGCTTTTGGCATTCGGCACTTCAGGAATATCCGCAACCGTTACCGCGCATACC 1820
Qy      601 LeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyLysThrValSerLeuAsn 620
Db      1821 CTGCAAGCCATGCTTTCAAGCGCAGCACTTCATCGCGGTAACCACTGTCAGCCTGAAT 1880
Qy      621 GlnAspLeuThrLeuGlnProTyrGlnValMetTrpLeuGluLeuAla 636
Db      1881 CAGGATTTGACCTTCAGCCCTATCAGGTCATGTCGCTCGCAATGCC 1928

RESULT 7
AY099335
LOCUS      1967 bp      DNA      linear      BCT 13-JAN-2003
DEFINITION      Neisseria polysaccharea strain 85322 amylosucrase Ams (ams) gene,
complete cds.
ACCESSION      AY099335
VERSION      AY099335.1 GI:27728139
KEYWORDS
SOURCE
ORGANISM      Neisseria polysaccharea
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE
AUTHORS      Zhu,P., Teang,R.S.W. and Teai,C.M.
TITLE      Nonencapsulated Neisseria meningitidis Strain Produces Amylopectin
from Sucrose: Altering the Concept for Differentiation between N.
meningitidis and N. polysaccharea
J. Clin. Microbiol. 41 (1), 273-278 (2003)
JOURNAL
PUBMED      12517860
AUTHORS      Zhu,P.
TITLE      Direct Submission
SUBMITTED (23-Apr-2002) LBP/DBPAP, CBSE, 8800 Rockville Pike,
Bethesda, MD 20892, USA
FEATURES
source
1..1967
/organism="Neisseria polysaccharea"
/mol_type="genomic DNA"
/strain="85322"
/db_xref="taxon:489"
21..1931
/gene="ams"
21..1931
/gene="ams"
/codon_start=1
/transl_table=11
/product="amylosucrase Ams"
/protein_id="AA051153.1"
/db_xref="GI:27728140"
/translation="MLTPTQVGLILQHLKTRILLDIYTPBOAGIEKSESDWRQPSRRM
DTHFPLKMLNSVYTGNEALLPMLLQAQWQISQASLKNKIDIERENPDWIL
SNKQGVVYDLDPFAGDLKDKIPYQELGLTYLHMLPLFKCPBGSKSDGAVSSY
RDVNPALGTDIDREVI AALHEAGISAVVDFTFNHTSNEHWAQRCAAGDPLFDFNY
IPDRMPQDYDTREIPPDQHPGPGSLDEGRWWTTFNSFDWLDLNSNPWFVRAM
AGEMLFLANLGVDIRMDAVAFWKMGTSCEMLPOAHALIRAFNAVNRIRAPAVFPK
SEAIYHPQVQVYIGQDSCQLYNPLQWALLANTLAREVNLHQAULTYRNLPEHTS
WNYVRSDDIGWTADADAAYLGISGVDRHQFLNRFVNRFDGSPFARGVFPNPNT
GCRVSGSAALAAQNDPRAVSRKLTISTALSTGGLPIYLDGVTGDPNDQWQV
DGKSDSDRWAHRPYNALYAQRNDPTAAGQIYQGLRHMIARVQSNRPFDFGGLVT
FNTNNKHIGYIRNALLAFNFSFPQVTAHTLQAMPFKAHDLIGKTVSLNQDLT
LPQVQWVLEIA"
459 a      631 c      489 g      388 t
BASE COUNT
ORIGIN

```

```

Alignment Scores:
Pred. No.:      1,95e-280      Length:      1967
Score:      3333.00      Matches:      620
Percent Similarity:      98.90%      Conservative:      9
Best Local Similarity:      97.48%      Mismatches:      7
Query Match:      97.80%      Indels:      0
DB:      1      Gaps:      0

US-09-843-007A-2 (1-636) x AY099335 (1-1967)

Qy      1 MetLeuThrProThrGlnGlnValGlyLeuIleLeuGlnTyrLeuLysThrArgIleLeu 20
Db      21 ATGTGACCCCCACGACGAAAGTCGGTTTGTATTTTACAGCACCTCAAACCGCATCTTG 80
Qy      21 AspIleTyrThrProGluGlnArgAlaGlyIleGluLysSerGluAspTyrArgGlnPhe 40
Db      81 GACATCTACAGCCCGAACACGCGCGCGCATCGAAAAATCCGAAGACTGCGCGCAGTTT 140
Qy      41 SerArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGly 60
Db      141 TCGCGCGCATGATGATGATTTCCCAAAATGTATGAAACGAACTCGACAGCGGTGTACGCG 200
Qy      61 AsnAsnGluAlaLeuLeuProMetLeuGluMetLeuAlaGlnAlaTyrGlnSerTyr 80
Db      201 AACACGAGGCCCTCTGCTGCTATGCTGGAATGCTCTGCGCGCAGCATGCGAGCTAT 260
Qy      81 SerGlnArgAsnSerSerSerLeuLysAspIleAlaArgGluAsnAsnProAspTyr 100
Db      261 TCCCAACGCGCGCGTCCCTGAAAAACATTCGACATCGAACCGCAAAACAAACCCGATTTG 320
Qy      101 IleLeuSerAsnLysGlnValGlyValCysTyrValAspLeuPheAlaGlyAspLeu 120
Db      321 ATTTGTCCCAACAAACAAAGTCGCGCGCGTGTGCTGCTGATTTGTTGTCGCGCGCATTTG 380
Qy      121 LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeu 140
Db      381 AAGGCTTGAAGACAAATCCCTTATTTTCAAGAGCTTGGTTTGACTTATCTGACCTG 440
Qy      141 MetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAlaValSerSerTyr 160
Db      441 ATGCGCTGTTTAAATGCCCTGAAGGCAAAAGCGCGCGCTATCGGTGACGAGCTAC 500
Qy      161 ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAlaLeu 180
Db      501 CGCATGTCAATCCGCACTGCGGCAATAGACGACTTCCGCAAGTATTGCCGCGCTG 560
Qy      181 HisGluAlaGlyIleSerAlaValAspPheIlePheAsnHisThrSerAsnGluHis 200
Db      561 CACGAAGCGGTATTTCGCGCTGCTGATTTCTATCTTCAACCAACACCTCCCAACGACAC 620
Qy      201 GluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePhe 220
Db      621 GAATGGGCGCAACGCTGCGCGCGCGGACCCCGCTTTTCGACAAATTTCTACTATATTTC 680
Qy      221 ProAspArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGln 240
Db      681 CCGACCGCGCATGTCGCGCAATACGACCGCACCTCGCGCAAAATCTTCCCGGACCG 740
Qy      241 HisProGlyGlyPheSerGlnLeuLysAspLysArgTyrValTyrThrPheAsnSer 260
Db      741 CACCGCGCGCTTCTCGCAACTGGAGACGAGCGCTGGGTGTGACGACGCTTCAATTC 800
Qy      261 PheGlnTrpAspLeuAsnTyrSerAsnProTyrValPheArgAlaMetAlaGlyLeuMet 280
Db      801 TTCCAATGGGACTTGATTAACAGCAACCCGCTGGGTATTTCGCGCAATGCGCGGCAATG 860
Qy      281 LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTyr 300
Db      861 CTGTTCCTTGCACCTTGGCGGTTCGACATCTTGGGTATGAGATGCGGTGCTTTATTTCG 920
Qy      301 LysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPhe 320
Db      921 AACAATGGGTACAGCTCGAANAACCTGCGCGAGGCGCACGCCCTCATCGCGCGTTC 980

```

QY 321 AsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHis 340  
Db 981 AATGCGCTTATGCGTATCGCGCGCGCGCGTCTTCAAAATCCGAAGCCATCGTCCAC 1040  
QY 341 ProAspGlnValValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnProLeu 360  
Db 1041 CCCGACCAAGTCTCCAAATATATCGGGCAGGAGCAATGCCAATCGCTACACCCCTTG 1100  
QY 361 GlnMetAlaLeuLeuTyrAsnThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAla 380  
Db 1101 CAAATGGCATTGTGTGGAACACCTTGGCCAGCGGAAAGTCAACCTGCTCCATCAGGCG 1160  
QY 381 LeuThrTyrArgHisAsnLeuProGluHisThrAlaTyrValAsnTyrValArgSerHis 400  
Db 1161 CTGACCTACCGCAACCACTGCGCGGAGCATCTCTGGGTCACTACCTGCGACCCAC 1220  
QY 401 AspAspIleGlyTyrThrPheAlaAspGluAspAlaTyrLeuGlyIleSerGlyTyr 420  
Db 1221 GACGACATCGGCTGGACGTTGCGGATGAAGATCGGCATATCTGGGCATAGAAGCGCTAC 1280  
QY 421 AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg 440  
Db 1281 GATCACCGCAATCTCTCAACCGCTTCTTGTCTCAACCGTTTGCAGCGGAGCTTGGCGGC 1340  
QY 441 GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla 460  
Db 1341 GCGGTACCTTCCAAATACACCCCAACACCGGGGACTGCGCGGTCAAGCGGTACGTCGCGC 1400  
QY 461 AlaLeuValGlyLeuAlaGlnAspAspProHisAlaValAspArgIleLeuLeuTyr 480  
Db 1401 GCATTCGCGGCTTGGCTCAAAATACCCCGCCAGCGCTGAGCGGCATCAAACTTTGTAC 1460  
QY 481 SerIleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr 500  
Db 1461 AGCATCGCATTGAGTACCGCGGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1520  
QY 501 LeuAsnAspAspTyrPheSerGlnAspSerAsnLysSerAspSerArgTyrAlaHis 520  
Db 1521 CCCAAGCAGCGCGCTGGCGGAGCAGCAACAAAGGAGCAGCAGCGCTGGCGGCAC 1580  
QY 521 ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGly 540  
Db 1581 CGTCCGCGCTACACAGCGCTTACGCGCAACGCAAGCATCGTCCGACCGCGCGG 1640  
QY 541 GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp 560  
Db 1641 CAAATCTATCAGGCTTGGCGCATATGATTCGCTGCGGCAAGCAATTCGCGCTTCGAC 1700  
QY 561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleGlyTyrIleArgAsn 580  
Db 1701 GCGCGCGGCTGTTACTTTCAACACCAACCAACACATCATCGCTACATCCGCAAC 1760  
QY 581 AsnAlaLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThr 600  
Db 1761 AACGCGCTTTGCGATTGCGCAACTTCAGCGAATATCCGCAACCGTTACCGCGCATACC 1820  
QY 601 LeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyValSerLeuAsn 620  
Db 1821 CTGCAAGCCATGCTTTTCAAGCGCGACGACCTCATCGCGGCAAACTGTCAGCTTGAAT 1880  
QY 621 GlnAspLeuThrLeuGlnProTyrGlnValMetTyrLeuGluIleAla 636  
Db 1881 CAGGATTTGACGCTTCAGCCCTATCAGTTCATGTCATGTCGCTCGAATCGCC 1928

RESULT 8  
A47369  
LOCUS 2883 bp DNA linear PAT 07-MAR-1997  
DEFINITION Sequence 1 from Patent WO9531553.  
ACCESSION A47369  
VERSION A47369.1 GI:2301369  
KEYWORDS Neisseria polysaccharea

ORGANISM Neisseria polysaccharea  
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
Neisseriaceae; Neisseria.  
REFERENCE 1 (bases 1 to 2883)  
AUTHORS Kosmann, J., Buettcher, V. and Welsh, T.  
TITLE DNA SEQUENCES CODING FOR ENZYMES CAPABLE OF FACILITATING THE  
SYNTHESIS OF LINEAR alpha -1,4 GLUCANS IN PLANTS, FUNGI AND  
MICROORGANISMS  
JOURNAL Patent: WO 9531553-A 1 23-NOV-1995;  
INST GENBIOLOGISCHE FORSCHUNG (DE)  
COMMENT Other publication DE 447388 960627  
Other publication AU 2614195 951205.  
FEATURES  
Location/Qualifiers  
1..2883  
/organism="Neisseria polysaccharea"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:489"  
/clone="PNE2"  
/clone\_lib="GENOMIC LIBRARY IN PBLUESCRIPTII SK"  
939..2783  
/note="unnamed protein product"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA03058.1"  
/db\_xref="GI:2301370"  
CDS  
661 a 852 c 758 g 612 c  
ORIGIN  
Alignment Scores:  
Pred. No.: 1,58e-263 Length: 2883  
Score: 3141.50 Matches: 621  
Percent Similarity: 98.11% Conservative: 3  
Best Local Similarity: 97.64% Mismatches: 11  
Query Match: 92.18% Indels: 9  
DB: 6 Gaps: 1  
US-09-843-007A-2 (1-636) x A47369 (1-2883)  
QY 1 MetLeuThrProThrGlnGlnValGlyLeuIleLeuGlnTyrLeuLysThrArgIleLeu 20  
Db 939 ATGTTGACCCCGCAGCAGTGGTGTGATTTTACAGTACCTCAAAACCGCATCTTG 998  
QY 21 AspIleTyrThrProGluGlnArgAlaGlyIleGlyLysSerGluAspTyrArgGlnPhe 40  
Db 999 GACATCTACACGCCGACAGCGCGCATCGAAAAATCCGAAGACTGCGCGCAGTTT 1058  
QY 41 SerArgArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGly 60  
Db 1059 TCGCGCGCGCATGGATACCGCATTTCCCAAACTGTATGAACGAACTCGACAGCGTGACCGC 1118  
QY 61 AsnAsnGluAlaLeuLeuProMetLeuGluMetLeuLeuAlaGlnAlaTyrGlnSerTyr 80  
Db 1119 AACAAAGAGCCCTGCTGCTATGCTGAAATGCTGCTGCGCGCAGGCATGCAAGCTAT 1178  
QY 81 SerGlnArgAsnSerSerLeuLysAspIleAlaArgGluAsnAsnProAspTyr 100  
Db 1179 TCCCAACGCAACTCATCTTAAAGATATCGATATCGCGCGCAAAACAAACCCGATGCG 1238  
QY 101 IleLeuSerAsnLysGlnValGlyValCysTyrValAspLeuPheAlaGlyAspLeu 120  
Db 1239 ATTTTGCCCAACAAACAGTCCGCGCGGTGTGCTACGTTGATTTGTTGCGCGCGATTG 1298

QY 121 LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeu 140  
 DB 1299 AAGGGCTTGAAAGATAAAATTCCTTATTTCAGAGCTTGGTTGACATTATCTGCACCTG 1358  
 QY 141 MetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyr 160  
 DB 1359 AFGCCGCTGTTTAAATGCCCTGAAGCAAAAGCAGCGCGGTATGCGGTACGACGATAC 1418  
 QY 161 ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAlaLeu 180  
 DB 1419 CGCGATGTCAATCCGGCACTGGGCACAATAGCGGACTTGGCGGAAGTCAATTGCTGGGCTG 1478  
 QY 181 HisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSerAsnGluHis 200  
 DB 1479 CACGAATGCG--AATTCCGCGCGTGTGCTGATTTATCTCAACCAACCACTCCCAACGAACAC 1536  
 QY 201 GluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAspPheTyrTyrIlePhe 220  
 DB 1537 GAATGGCG-CAACGCTGC---GCCGCGCAGCCGCTTTCGACAATTCCTACTATATTTC 1592  
 QY 221 ProAspArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGln 240  
 DB 1593 CCCGACCGCGGATGCCGACCAATACGACCGCAACCTCGCGGAATCTTCCCGACGACG 1652  
 QY 241 HisProGlyGlyPheSerGlnLeuGluAspGlyArgTyrValTyrThrPheAsnSer 260  
 DB 1653 CACCCGCGCGCTTCTCGCACTGGAAGACGACCTGGGTGTGACGACCTTCAATTCC 1712  
 QY 261 PheGlnTrpAspLeuAsnTyrSerAsnProTyrValPheArgAlaMetAlaGlyGluMet 280  
 DB 1713 TTCCAATGGGACTTGAAATTCAGCAACCCGCTGGGTATTTCG- GCAATGGCGGGGAAATG 1771  
 QY 281 LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTrp 300  
 DB 1772 CTGTTCCTTGCCAACTGGGCGGTGACATCTCGGTATGGAATGGGTTCCTTTATTTCG 1831  
 QY 301 LysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPhe 320  
 DB 1832 AAACAAATGGGACAGCTGCAAACTGCG--GCAGCGCAGCGCTCATCGCGGCTTC 1889  
 QY 321 AsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHis 340  
 DB 1890 AATGCGGTATGCGTATTGCGCGCGCGCGCTGTTCTTCAATCCGAAGCCATCGTCCAC 1949  
 QY 341 ProAspGlnValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnProLeu 360  
 DB 1950 CCCGACCAAGTCTGCTCAATACATCGGCGAGGACGAATGCCAAATCGGTACACCCCTG 2009  
 QY 361 GlnMetAlaLeuLeuTrpAsnThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAla 380  
 DB 2010 CAATATGGCATTTGTTGGAAACACCCCTTGCCACGCGCGAAGTCAACCTGCTCCAACGCG 2069  
 QY 381 LeuThrTyrArgHisAsnLeuProGluHisThrAlaTrpValAsnTyrValArgSerHis 400  
 DB 2070 CTGACCTACCGCCCAACCTGCGCGAGCATACCGCTTGGGTCAACTAGTCCGCGACGCC 2129  
 QY 401 AspAspIleGlyTyrTrpPheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyr 420  
 DB 2130 GACGACATCGCTGGAGCTTTGCGCATGACAGCGCGGCATATCTGGGCATAAGCGGCTAC 2189  
 QY 421 AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg 440  
 DB 2190 GACCACCGCCCAATTCCTCAACCGCTTCTTCGTCAACCGTTTCGACGCGCACGCTTCGCT 2249  
 QY 441 GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla 460  
 DB 2250 GCGCTACCGTTCCTCAATACACCAAGCACAGCGGACTGCCGTGTGCTGATGACGCGCG 2309  
 QY 461 AlaLeuValGlyLeuAlaGlnAspAspProHisAlaValAspArgIleLysLeuLeuTyr 480  
 DB 2310 GCATTGTCGCTGGCGCAAGCAGCATCCCGACGCGGTGACCGCATCAAACTCTGCTAC 2369  
 QY 481 SerIleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr 500

DB 2370 AGCATTTGCTTTAGTACCGCGGCTCTCCGCTGATTTACCTAGCGACGAATGGGTACG 2429  
 QY 501 LeuAsnAspAspAspTyrSerGlnAspSerAsnLysSerAspAspSerArgTyrAlaHis 520  
 DB 2430 CTCATGACGACGACTGCTGTCGAAGC-AGCAATAGAGCGGACGACGCGGTGGGC-CAC 2487  
 QY 521 ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGly 540  
 DB 2488 CGTCGCGCTACAAACGAAGCCCTGTACGCGCAACCGAAGCATCGTCGACCGGACGCGC 2547  
 QY 541 GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp 560  
 DB 2548 AA-ATCTATCAGGGCTTTCGCCCATATGATTCGCTCGCCAAAGCAATCCCGCTTCAC 2606  
 QY 561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArgAsn 580  
 DB 2607 GCGCGCAGGCTGGTTACATTCACCAACCAACCAACCAATCATCGCTACAT-CCGAAC 2665  
 QY 581 AsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThr 600  
 DB 2666 AATCGCTTTTGGCATTCGGTAACCTTCAGCGAATATCCGCAACCGTTACCGCGCATACC 2725  
 QY 601 LeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyLysThrValSerLeuAsn 620  
 DB 2726 CTGCAAGCCATGCGCTTCAGGCGCAGCACTCATCGGTGCAAAACTGTTCAGCGCTGAT 2785  
 QY 621 GlnAspLeuThrLeuGlnProTyrGlnValMetTrpLeuGluIleAla 636  
 DB 2786 CAGGATTTGAGCTTCAGCCCTATCAGCTCATGTGCTCGAATTCGCGC 2833  
 RESULT 9  
 A47785 LOCUS A47785 2883 bp DNA linear PAT 07-MAR-1997  
 DEFINITION Sequence 1 from Patent WO9533838.  
 ACCESSION A47785  
 VERSION A47785.1 GI:2301671  
 KEYWORDS Neisseria polysacchara  
 SOURCE Neisseria polysacchara  
 ORGANISM Neisseria polysacchara  
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 Neisseriaceae; Neisseria.  
 Kossman,J., Buetcher,V. and Welsh,T.  
 MICROORGANISMS PERMITTING THE INTRACELLULAR POLYHYDROXY ALKANATE  
 SYNTHESIS WITH SIMULTANEOUS EXTRACELLULAR POLYSACCHARIDE SYNTHESIS  
 AND PROCESSES FOR PRODUCING THE SAME  
 Patent: WO 953383-A 1 14-DEC-1995;  
 INST GENBIOLOGISCHE FORSCHUNG (DF)  
 Other publication AU 2787895 960104  
 Other publication DE 4420223 950504.  
 FEATURES  
 source  
 1. .2883  
 /organism="Neisseria polysacchara"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:489"  
 /clone="PNE2"  
 /clone\_lib="GENOMIC LIBRARY IN PBLUESCRIPTII SK"  
 939, 2783  
 /note="unnamed protein product"  
 /codon\_start=1  
 /translation="MLTPTQVGLILQYLKTRILDIYTPQAGIKESDWRQFSRRM  
 DTHPKVNLDSVYGNNEALPMLEMLAQWQSYORSNLSKDDIDIAENPNPWTL  
 SNKQVGVYDLFAGDLGKDKIYFQELGTLLEMLPKFCPSKSGGYAVSYT  
 RYNPALGTGLDLREVIALLHSHFRFRFYLPQLQRTMAORCAGDPLDFNYLIF  
 EDPMQDQDRLRIFFPDHPGFGSLEDGRWVWTFNSFQMDLNTSNPWFAQWRA  
 KCCSLTWTALTCVWRLPLGNKGQAATCAHALIRAFNAWMLRAAPVFKSEA  
 IVPDQVQVIGQDEQIGVNPLOQWALNTLREVLHLYALTYRHLPEHTAWN  
 YVSHDDIGWTFADDEAAVYLGISYDHRQLEFPVNRFDGTGARGVPQFNPSTGDC  
 RVSGTAALVGLAQDDPHAVDRILKLYISALSTALSTGLPLIVLGDVGTLLNDQCAAI



DEFINITION Deinococcus radiodurans R1 section 83 of 229 of the complete chromosome 1.  
ACCESSION AE001946  
VERSION AE001946.1  
KEYWORDS GI:6458655  
SOURCE Deinococcus radiodurans  
ORGANISM Deinococcus radiodurans  
Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales; Deinococcaceae; Deinococcus.  
REFERENCE 1 (bases 1 to 10719)  
White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L., Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C., Makarova, K.S., Aravind, L., Daly, M.J., Fraser, C.M. et al.  
Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1  
Science 286 (5444), 1571-1577 (1999)  
JOURNAL 20036896  
MEDLINE 10567266  
PUBMED  
REFERENCE 2 (bases 1 to 10719)  
White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L., Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C., Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W., Fleischnann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S., Smith, H.O., Venter, J.C. and Fraser, C.M.  
Direct Submission  
Submitted (08-NOV-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
FEATURES  
source  
1..10719  
/organism="Deinococcus radiodurans"  
/mol\_type="Genomic DNA"  
/strain="R1"  
/db\_xref="taxon:1299"  
/chromosome="1"  
72..1034  
/gene="DR0932"  
72..1034  
/gene="DR0932"  
/note="similar to SP:P19641 GB:X68873 PID:388220  
PID:606125 GB:U00096 percent identity: 57.98; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="polyphenyl synthase"  
/protein\_id="AAFI0509.1"  
/db\_xref="GI:6458656"  
/translation="MPDAAFETRMREVLRSDFEITAGIEDIVTAGKKRVLPLVTLA  
AALGAGDPAATSHSWDELRLSYCVLLHSALLHDDLDLDDSKRGQVTAFFPG  
NVVYMSGDGEMALRLGLSGMPSGPAITRAFGQAASVICBGEVLQFOVASADYSFE  
NYFQVHGKTAALLRLAQAAPMLGADDSARDALSTGRYGHAFQMDLLDLLE  
EAQIKGVGGDREGKATYFVLCGGGPHDAEVALERRAGSDGVARMALAKQER  
TDDCTREIRRRRLAIALDALPSPARTALAAERLARNTH"  
1109..3043  
/gene="DR0933"  
1109..3043  
/gene="DR0933"  
/note="similar to PID:2094767 percent identity: 60.15;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="alpha-amylase"  
/protein\_id="AAFI0510.1"  
/db\_xref="GI:6458657"  
/translation="MLTFDLAARLAFDDDRDAETFRRLRLRYGPELADNLRAVYGN  
RADALIGELVLMHAYHAPADLRDLERALLRDLWDQSGEMVGVYVDRPAGTLR  
VGCEELVGLGVLYLHLLRDLRDEGGVAVQDYSVRPDLGTFIDLSALARE  
LRGRIEVLIVLVAEHEHVAATGAAYRDYHIFDRTPQDPAYRTLPEIF  
PDFAPGNTFTEAGGWTTFNRSQMDVWGNPAVFRFYLIDLTILNARVGEVRLD  
AIALFWKELGTDCQNPQVHRLTHALRAATRIVAPAVAKAAIVAPGDLIHYLGRD

HHGRVSDWAYHNSLMVQWSSLASRDTLLTAALAAFPKPTNTTGWVYVRCHDIGW  
AIADDAARVGLSGPAHRLFLSDYSGFFPGSGFARGLVFQHHPOGQDRISGTASLA  
GLDLALETGDAERVNDALAILLHVMVGFPVLLVMGDELALNDDTDAFAVAHA  
ADNRVHEPQMDWELVASAQADAATQGPVTPAGMPAGLHLAVERRTPHLHASTES  
PLPSPPDQVLLIRRHPTGVLLQVYNFSEHHTFPTWPLQEQLGAVAHDLLEGSQFHF  
LGGDLALEPFRALWLVAGG"  
3065..3838  
/gene="DR0934"  
3065..3838  
/note="similar to GB:AE000657 percent identity: 72.64;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="transcriptional regulator, MerR family"  
/protein\_id="AAFI0511.1"  
/db\_xref="GI:6458658"  
/translation="MLVVCSTSLMGSDAKHEPVYVISVAABLVDMPHQPOTLRYERKGL  
IRPGRSGKTRLYSERDIEHLREIRRLTQELGVXAGVEVRLRQHQDDDDQQEFAE  
IERIEDELRQAPRALPAPDPDPDRPVVISIAAABLVDMPHQPOTLRYERKGLIH  
PGRSGKTRLYSERDIEHLREIRRLTQELGVXAGVEVRLRQHQDDDDQQEFAE  
FLQDDLSERMTTWTLPAPSEAGPEEAAAADAGEDQN"  
3816..4907  
/gene="DR0935"  
3816..4907  
/note="similar to GB:D10483 SP:P05637 GB:X04711 PID:216474  
PID:40919 percent identity: 55.48; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="phosphatase, putative"  
/protein\_id="AAFI0512.1"  
/db\_xref="GI:6458659"  
/translation="MPEKTRTEHQELSLTFLRGVRLRREPRFVSDALSRPSAAPPVIL  
CPEVFLNALALPLSPVPAFLWVGVDVGRGADKRLALLAGLTHGHTWGAHLH  
VFLGDDYDRGPDGLGVRLVQSLREQLAGCEVTALLGNHEVMLLAALFREQDPRD  
RFGFFDYRNSGQGLTDLERLAPGLDLARPLLAHAQGWLLHADSPPFVLGNDP  
QEVNRYAEILLSPDPGMPAFAFVLDLKGAGDCTQARLLETGYGGERLVHGT  
FTPLLQSGQDLPGEFTFVLYHADGLCLAVDSGHAYFPDAGFVARLGBELAPGGAER  
VLAELVLPFPFLDGLSLTALRPSGEG"  
4977..6017  
/gene="DR0936"  
4977..6017  
/note="similar to GB:AL009126 percent identity: 61.90;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAFI0513.1"  
/db\_xref="GI:6458660"  
/translation="MPILLDLHPDHPLEGRKOLLEWVVGVTGTPDAMTNLPAGL  
RAELSYQLNPREDIETVSHSGSVKYLITLNDGRQMEAVMYPLRKTICVSTMVG  
KAPCAFCATGKMGFGNLTGPEITGVQVLAVAGGEGISPREINLVFMGMEALLNYE  
NSMAARLHPDGLMGSKERTVLTSTVIAGKIRLAEEDDLGKLSLHAPDEETE  
QQIIPGTANSIEIIMAAADYQAVTGRVYTMELRGVNDHLWAEILLADRLGLV  
SHVNLIPMNWDGSDFVSSSEEQIQAIFYDALQGRGVDSVRSRSGDKAGAACGQLAR  
RPGATGEVGA"  
6345..8135  
/gene="DR0937"  
6345..8135  
/note="similar to PID:1652192 percent identity: 41.71;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="tetrapeptide repeat family protein"  
/protein\_id="AAFI0514.1"  
/db\_xref="GI:6458661"

```

NEFEALIVRNYPNPEPHFGLGLALYQLGLRGATPFEGQMFPAPQRYEGPYNLGVIA
TRGRYEDALRLYEALTLQAQSPAPAAQVLLRALATEQGRVXYVALAATYADLRRA
DPENVEYIFRHAQLYQGVADALPVYAVLERKFSLQAALLADLYVAQGLPORA
VRELSAAARVLCTGTATLLRQSDLLAQNRLGALNAASEARDRLPRAIIRE
AELLAQCGQPPAAVRYDALALTPKAPITQAALALYLGNOYPERAAQAOALKIT
PDASTARALYQVAVRYGOVTOARTALNSSVAVPDSLLMLGLSLYALKDVAS
APVLSVESVLRPTPARONLAALATAYAEALRLGLVTDEVENAGWYLLGLS
ORAQORFDDARQFKTAAGLSARAKDALK"
/gene .9319
/gene="DR0938"
/gene="DR0938"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAFI0516.1"
/db_xref="GI:6458663"
/transl_table="MTSARTPSRPRPRPAGRPPELVAATLVLAGVGLGS
LLGERAPAPVCAQOPTTIPVSGAVADGSPOTAAQAQCTAGTKIPAPAAP
AKTIPPVLPPEPRMPAPPTPEVEVETTTAASPIQPPVTOAPATQTPPTPRAQA
PATQPPATPAPAPAPAPATPTEPATTEPAAPAGTPAVAPASQRTPLASDRV
MLGTFGERALATAGVAGLGVTVYVIDUGNOFVAGQFPFADEASGQQAADIRAY
ARAEYLFVEGOTLUNPAPAAAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP
FNNVEGAQRFEQLRAQGFSPVNAPEYTKVTVLLGLPLTGSLDSTEGRLDAAGLDHF
RLR"
/gene .10020
/gene="DR0939"
/gene="DR0939"
/notes="similar to GB:AL009126 percent identity: 51.93;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAFI0515.1"
/db_xref="GI:6458662"
/transl_table="MSRVADIPATVGRVVYTVIRVLEELQAQVLRASSGELARRAGV
TPFQVKKDLTFGRGTRGIGYTVAVLERELRLALGLDQTNVTVVMGRGLHAIANY
PGASDYOFQNVGFDVAPDVGVREVRGLTIQHSQGPFPVAVAGTPROVDMGLTVP
AEHAQAAQALVAGVGGILNFPVLTQDLHLPEAFAPGRREVTVENVDFLAGWK
RLAFYMLGPQAGPAES"
/gene .10017
/gene="DR0940"
/gene="DR0940"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAFI0517.1"
/db_xref="GI:6458664"
/transl_table="MTWKMPFPVLLGLLAGCGTGVGLSGRVNVGVVDGADAGSQV

```

## Alignment Scores:

```

Pred. No.: 4,35e-104 Length: 10719
Score: 1319.50 Matches: 281
Percent Similarity: 58.62% Conservative: 110
Best Local Similarity: 42.13% Mismatches: 215
Query Match: 38.72% Indels: 61
DB: 1 Gaps: 14

```

US-09-843-007A-2 (1-636) x AE001946 (1-10719)

```

QY 1 MetLeuThrProThrGlnGlnValGlyLeuLeuGlnTyrLeuLysThrArgGlyLeu 20
Db 1109 ATGCTACGCTGAC-----CTCGCCGCCGCCCTG--- 1138
QY 21 AspIleTyrThrProGluGlnArgAlaGlyIleGluLysSerGluAspTTPArgGlnPhe 40
Db 1139 -----CGCTCGCTTTCAGCATGACCGTGAAGCGCGAGACTTTC 1177
QY 41 SerArgMetAspThrHisPheProLysLeuVecAsnGluLeuAspSerValTyrGly 60

```

```

Db 1178 CGCTCGCGCTGGAGCGGTACGCGCGGAACCTGGCGGCAACCTCGGCGGTCTACGGA 1237
QY 61 AsnAsn---GluAlaLeuLeu---ProMetLeuGluMetLeuLeuAlaGlnAlaTrpGln 78
Db 1238 AACACGCGACGCGCTGATAGCGAGCTTCTAGAGTCACTG-----CAC 1285
QY 79 SerTyrSerGlnArgAsnSerSerLeuLysAspIleAspIleAlaArgGluAsnAsnPro 98
Db 1286 GCTTACACGCGCGCGCGCGCGACCTCAACACGCTGACGAGCGCGGTGCTCGCGCCC 1345
QY 99 AspTyrIleLeuSerAsnLysGlnValGlyGlyValCysTyrValAspLeuPheAlaGly 118
Db 1346 GACTGGCTCAAGGACCGAGATGGTGGCTACGTGGCCCTAGCTGACCGCTTGGCGGC 1405
QY 119 AspLeuLysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuTyrLeu 138
Db 1406 AGCTTCAGGCGGTGGCGAGCGGTGGAGTATCTGGAAGGGTTGGCGCGTACCTT 1465
QY 139 HisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSer 158
Db 1466 CACTCTCTCGCTGCTCGGCGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1525
QY 159 SerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAla 178
Db 1526 GATTACCGGAGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1585
QY 179 AlaLeuHisGluAlaGlyIleSerAlaValAspPheIlePheAsnHisThrSerAsn 198
Db 1586 GAGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 1645
QY 199 GluHisGluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyr 218
Db 1646 GAACACGAATGGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1705
QY 219 IlePheProAspArgMetProAspGlnTyrAspArgThrLeuArgGluLeuPhePro 238
Db 1706 ATTTTCTCTGATCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1765
QY 239 AspGlnHisProGlyGlyPheSer---GlnLeuGluAspGlyArgTyrValTyrThr 257
Db 1766 GACTTCGCGCGGTGAATTTTCACTGGAACGGAAGCGCGCGCGCGCGCGCGCGCGCGCG 1825
QY 258 PheAsnSerPheGlnTrpAspLeuAsnTyrSerAsnProTyrValPheArgAlaMeta 277
Db 1826 TTCAACCGCTCCAGTGGGACGTGAACCTGGGGAACCCCGCGCGGTGTTCGCGGATATCTC 1885
QY 278 GlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAla 297
Db 1886 GACTCATTTCTCACTCTCGCAACCGGCGCGGTGGAGGTCTTCGCGCTCAGCCATCGCT 1945
QY 298 PheIleTyrLysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIle 317
Db 1946 TTTTGTGGAAGCGGTGGGCGACCGACTGTCAAGAACCCAGCTTCAGCGCTCCACCGCTGACC 2005
QY 318 ArgAlaPheAsnAlaValMetArgIleAlaProAlaValPhePheLysSerGluAla 337
Db 2006 CACGCGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2065
QY 338 IleValHisProAspGlnValGlnTyrIleGlyGlnAspGlu----- 352
Db 2066 ATTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTC 2125
QY 353 CysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTyrAsnThrLeuAlaThrArg 372
Db 2126 TCCGACATGGCTTACCAACCTCGCTGATGGTGGAGCTCTGGAGAGCGCTCGCTCGCGGT 2185
QY 373 GluValAsnLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHisThrAla 392
Db 2186 GACACGCGGTCTCTCACGCGCGCGCTGGCGCGCTTCCCCCAACCGACGACGACGAC 2245
QY 393 TrpValAsnTyrValArgSerHisAspIleGlyTyrThrPheAlaAspGluAspAla 412

```

Db 2246 TGGGGCGCTCTACGCGCTCTGTCACGACATCGCTGGCGGATCGCGCGACGAGCGCG 2305

Qy 413 AlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsnArgPhePheValAsn 432

Db 2306 GGGCGCTGGGCGCTCTACGCGACCGCGCGACCGCCACTCTCTTCGGGACTTTACAGCGGC 2365

Qy 433 ArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsnProSerThrGlyAsp 452

Db 2366 GAGTTTCCCGCGCTCTTTGGCGGGGACTGGTGTTCAGCACCCACACGACCGGGGAC 2425

Qy 453 CysArgValSerGlyThrAlaAlaLeuValGlyLeu----- 465

Db 2426 CGCGCGATTCGGCGACCGCGCGGAGCTTCGCGGGCTGGACCTCGCGTGGAAACCGGA 2485

Qy 466 ---AlaGlnAspAspProHisAlaValAspArgIleLysLeuTyrSerIleAlaLeu 484

Db 2486 GACCGCGAGCGGGTGAACGACGCGCTCGCGCGCTCTCTCTCTCGACGGGTCTATGCTC 2545

Qy 485 SerThrGlyGlyLeuProLeuLeuTyrLeuGlyValGlyThrLeuAsnAspAsp 504

Db 2546 GCCTTCGCGGGGCTCGCTCTCTGATGATGGCGACGACCTCGGTTCAACCGCGCGACATG 2605

Qy 505 AspTrpSerGlnAspSerAsnLysSerAspSerArgThrAlaHisArgProArgTyr 524

Db 2606 GACTTTCGCGCGCTCGCGCGCCACCGCGCGACCAACCGCTGGTTCAACCGCGCGACATG 2665

Qy 525 AsnGluAlaLeuTyrAlaGln-----ArgAsnAspProSerThrAla 538

Db 2666 GACGCGGAGCTGGTGGCTCTGCGGCGCGACCGCGCGACCGCGCGCGTCAACCGCC 2725

Qy 539 AlaGlyGlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArg 558

Db 2726 GCGCGCGAAGTTTCGCGGGCTCGACACCTCTGCGCGGTGCGCGCGCGACCCCGCAC 2785

Qy 559 PheAspGlyArgLeuValThrPheAsnThrAsnAsnLysHisIle----- 574

Db 2786 CTGCAACGCC-----AGCACCGAGTCTGCTCCCTCCCGAGTCCCGAC 2827

Qy 575 -----IleGlyTyrIleArg-----AsnAsnAlaLeuLeuAlaPheGlyAsn 588

Db 2828 CCTTCGCTGCTGCTGCTGCGCGCGACCGCGCGCGTCTGCTGCTGCTGCTGCTGCTGCT 2887

Qy 589 PheSerGlyTyrProGlnThrValThrAlaHisThrLeuGln---AlaMetProPheLys 607

Db 2888 TTCAGCGAGCACCATCATCTTCCCGACGCTGCGCGCTTGCAGGAACAGCTCGCGCGCGCTG 2947

Qy 608 AlaHisAspLeuIleGlyLysThrValSerLeu---AsnGlnAspLeuThrLeuGln 626

Db 2948 GCCACGACCGCTCGCTGAGAGCGCAGTTCACCTGGCGCGCGCGCGACCTCGCGCTCGAA 3007

Qy 627 ProTyrGlnValMetTrpLeu 633

Db 3008 CGTACCGCGCGCTGCTGCTG 3028

RESULT 11

AE011997

LOCUS

DEFINITION

Xanthomonas axonopodis pv. citri str. 306, linear BCT 29-MAY-2002

the complete genome.

ACCESSION

AE011997 AE008923

VERSION

AE011997.1 GI:21109853

KEYWORDS

SOURCE

ORGANISM

Xanthomonas axonopodis pv. citri str. 306

Xanthomonas axonopodis pv. citri str. 306

Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;

Xanthomonadaceae; Xanthomonas.

1 (bases 1 to 10261)

da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Parah, C.S., Furlan, L.R.,

Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida

Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,

Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardozo, J.,

Chambergo, F., Ciapina, L.P., Cicarelli, R.M.B., Coutinho, L.L.,

Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S.,

FERREIRA, R.C.C., FERRO, M.I.T., FORMIGHIERI, E.F., FRANCO, M.C.,

GREGGIO, C.C., GRUBER, A., KATSUYAMA, A.M., KISHI, L.T., LEITE

Jr., R.P., LEMOS, E.G.M., LEMOS, M.V.F., LOCALI, E.C., MACHADO, M.A.,

MADERA, A.M.B.N., MARTINEZ-ROSSI, N.M., MARTINS, E.C., MEIDANIS, J.,

MEICK, C.F.M., MIYAKI, C.Y., MOON, D.H., MOREIRA, L.M., NOVO, M.T.M.,

OKURA, V.K., OLIVEIRA, M.C., OLIVEIRA, V.R., PEREIRA, J.R.H.A.,

ROSSI, A., SENA, J.A.D., SILVA, C. de Souza, R.F., SPINOLA, L.A.F.,

TAKITA, M.A., TAMURA, R.E., TEIXEIRA, E.C., TEZZA, R.I.D., TRINDADE DOS

SANTOS, M., TRUFFI, D., TSAI, S.M., WHITE, F.F., SETUBAL, J.C. and

Kitajima, J.P.

Comparison of the genomes of two Xanthomonas pathogens with

differing host specificities

Nature 417 (6887), 459-463 (2002)

2202145

12024217

2 (bases 1 to 10261)

da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Parah, C.S., Furlan, L.R.,

Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida

Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,

Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardozo, J.,

Chambergo, F., Ciapina, L.P., Cicarelli, R.M.B., Coutinho, L.L.,

Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S.,

Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,

Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite

Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,

Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,

Meick, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,

Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira, J.R.H.A.,

Rossi, A., Sena, J.A.D., Silva, C. de Souza, R.F., Spinola, L.A.F.,

Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos

Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and

Kitajima, J.P.

Direct Submission

Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de

Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,

Brazil

Location/Qualifiers

1. 10261

/organism="Xanthomonas axonopodis pv. citri str. 306"

/mol\_type="genomic DNA"

/strain="306"

/db\_xref="taxon:190486"

/notes="pathovar: citri"

155. 2629

/genes="fyua"

/notes="synonym: XAC3489"

155. 2629

/genes="fyua"

/notes="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer/Genemark"

/codon\_start=1

/transl\_table=11

/product="TonB-dependent receptor"

/protein\_id="AA038332.1"

/db\_xref="GI:21109853"

/translation="MRERIAMSTLHTMELHALACAVVSCICAPALAAQDTAARAPATP

PAADSAVNDVSVFTGISTATITKXSSVSSTVGAEAQSPRSTAEIFRNPGRIR

SESSGEGNANIAVRGLPVASGAKTLQEDGLPVNFGDIAPFGIDNTRIFNYGAPFA

RIRAIKGSASTTINAPGGIINFISKTGTAGSGVSGRLDYDNTRIFNYGAPFA

EQVTFNIGGSGDQVDRDAGYTTDKGQLKXNLTRLFHNYGVYKGLNDRAAGYL

PWPTSVRGDQSPDLGCPGDPGNDTLVSRNFTDVLCDNQPRETDLGDGMHPSL

RTIGARAWFDLGNALRTHLENVAINDLGNVTDLVSVEFGDGTFLNRGYSSTOT

IDMDTWNSTVQSLGRNRLNVDAAGVSRQNGLYAYGTFPFGDCCVTRSDRYVD

VNAPYVATFDTSKPSIDGSLRYDMGDARSGTATAQNLVDNGDGVIVQVEORVAT

VTANSPRPVDYNLYLSYLGNGYLTIDLGAFARISGRANARADLFLFVGDDGVS

SSYRGVNVVQAENGLWRDGLSLFATFSARTQENFEITSQFFNRGDMYQVGL

EASVYRGFTLNGLTWDAEISRDQITPENAGNVPRQADVVMQLTPSVYRGDMYQVGL

VNIIGTDTATQANQLKMPGYTCVNLFGDVRVTDALTVALNVNLTFTGLTEAEEA

TIPIANGIIRARSAGRTTISLRYDF"

2664. 4598

/gene="XAC3490"

2664. 4598

gene

CDS



```

/gene="XAC3490"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="amylase or alpha amylase"
/protein_id="AAM38333.1"
/db_xref="GI:21109855"
/translation="MSTCIPDSPALRAFAAPLDPQAEVLLSRYSQHASRLDLALHA
LYGRADVASLWAGLQVGLVYRFLRAGAGDGFVSDYGOVPSLGSNDLVA
GTEQVAVSRVYDELGVYRFLRAGAGDGFVSDYGOVPSLGSNDLVA
LTSRLRAGISLADFLVHTADDEAQAARAGDARYLDYHFPADRLPDRYEATL
QVFPHTAPAGNTVDVDAOMMTTFYPYOWDLNPNPVAQVDMALAMKLNGLVET
FLDSTAYLWKRIGTDCWNOPEATLVALRAVTDIAPAVVMKAAIPEMTOLPVP
GSDVGEHECHLAHSLIMLAGSALALQCGDILHNVIAHSPLPCHCAWLSVXCHD
DIGNVLOHEACGNAAPPSRLRVARFANAPGVSARGESQSSGDGVHGTNGMAA
ALAGIQAQAGADALAAVADRLVLYATALAMFVGLIYMGDELAWMPGYRHDP
HQEGHRLHPAMDQAAVADRLVLYATALAMFVGLIYMGDELAWMPGYRHDP
ALNPRVFLTRGDSFTALHNSQPLDVELAAGIDGWTLLAIDDAIGGAAGGDS
IVLPYGVNLRQGTETAPE"
/complement(4772..5452)
/gene="nonP"
/notes="synonym: XAC3491"
/complement(4772..5452)
/gene="nonP"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="NonP-related protein"
/protein_id="AAM38334.1"
/db_xref="GI:21109856"
/translation="MKVLMVLTSHDQLDGTGKTGFWLEFAAPYVYFDKAGADITLV
SPRQGLDPAKSPDEPAQDTRFKFDQPEAQALASTHRLADVAEDFVLVFPGG
HGFUQDLAEDSIALLEPAEADKPGLVHAPGALRKYKSGDGTPLNGRRVGTFT
NSEEGVGLTKIVPLVEDVLTGLGKYERKAGWGVVYVDTGTLVGTQNPASSEKTA
TLALAKK"
/complement(5741..7003)
/gene="XAC3492"
/complement(5741..7003)
/gene="XAC3493"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAM38335.1"
/db_xref="GI:21109857"
/translation="MPSRKPPDIVHHAALDARLVKAVRGVRLALASMPRLQEPFL
QSVARGQPELPQVSEYITLDFGTRRELAAIAKADPHPIGEYLIDSNHWSLAAGLL
ESLGTAAVSDYSAGLFQVDPDMPGNGPSTREAAQHFIQIAELDELMAPEQVPS
ATLRMLQDLNDGDFEGRIITVTLPDLAKAAGAHRIKLTGATFSDYDRAQLFH
HEALVHSLTALNGRAQVHPLSLGTSPPRTATQGLATPAEQITGSDITERMKRISLR
IEAAMARSGADFTVEVRYTAGQSAESFSSAORVFRGPTSGGGAFTKDTVYLRG
LVAVTTPFRHSIORDLQCLCFAGKVALGDVAREAPLFESEVIVAPRPLPWVSRV
SGLAGMLAFSLFANEIMDQLQPPANNV"
/complement(7003..8094)
/gene="XAC3493"
/complement(7003..8094)
/gene="XAC3493"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="integral membrane nucleotide protein"
/protein_id="AAM38336.1"
/db_xref="GI:21109858"
/translation="MNSTSPDPTLQISGMFWLPEEAAQVFAAHPOALPARVVE
QHRGYVAVDTPASVAKESLPEWQRPRFPSEHRAAVGDMVLMGRIIVALLPRTSI
KRGAGHHVQVQVIAIDTVFVCGLDADNPRIERYLLVLLVGGGAQVPPVVLKAD
QTEVAREALAVLEEAONTPLRAVNAKOPDSVAALRPMLGDGRITAVLVSSGAKST
LNTVLLTGOKXNAUPENDSRGHTTHRALIPLSGACLDITDPMRELKPTGEDL
AEGGSVSEALAAQCRNDCHIAEFGCAVRAIDAGLDPEFVANTMKLENEVASAR

```

```

EXLATRVQNNRKGSGKRPASVDORPCRH"
/247..9518
/gene="aspC"
/notes="synonym: XAC3494"
/247..9518
/gene="aspC"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="aminotransferase"
/protein_id="AAM38337.1"
/db_xref="GI:21109859"
/translation="MSTAPQKPLAIRBELSERVIEIRGELARRARELEAOGRLIKLN
IGNRQAGFRAPHELQRAIADDMGRTDPTHQOGLPEAREAIATAYARRQHPDADPR
IPVNGVELLIDLSIRALLNPGDEVLPDLSAAATINDGRPYVYRCAPENGFO
PDEVEITVSSRTRAIVLINNNPNSGVSRELLERIVAIATKHNLMLVDEIYDQV
LYDDAAFPVAPLAGAPFCITFSGLSKVRACGWRVGLWALLSGDQARINDLRAMDLL
GALLCANVQQAIDAIVNGPDTISALCAFGGRLYETRAVTEACAAASHLSLVAFA
MALYAFADVAGAAARNFDDHAFALDLNVEBGLVVPGSSFNVPYRHHFRVTLMPGAV
MRDVPARIDRALARRAEAVTKVPLKPSVA"
/9518..10195
/gene="XAC3495"
/9518..10195
/gene="XAC3495"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAM38338.1"
/db_xref="GI:21109860"
/translation="MAAPGTPLRYALGDSYTTGEGVAPQAPRNPSPQLVDGLRAYGNV
APPOIATGTTWTDLQADAAAPQAPFELVLLGVNNYGRSLDRAVRSQFDALL
ORALGPAGAGAAVFLVSIPIWGLTPEFAHQGGDAALIGAIQIDAFNNVABRCAARSV
RFVDITATSRGSDADMLVDGLHPSGMYARWATLPAARDAATAANAIGWEHAM
RQGPSPSR"

```

## Alignment Scores:

Pred. No.:	2.6e-87	Length:	10261
Score:	1126.50	Matches:	240
Percent Similarity:	56.33%	Conservative:	107
Best Local Similarity:	38.96%	Mismatches:	236
Query Match:	33.05%	Indels:	33
DB:	1	Gaps:	10

US-09-843-007A-2 (1-636) x AE011997 (1-10261)

QY	43	ArgMetAspThrHisPheProLeuMetAenGluLeuAspSerValTyrGlyAsnAsn	62
DB	2751	CGCTACGACCAACGCGATCGCGTTCTGCGATGCGCTCTACGCGCCAGCGT	2810
QY	63	GluAlaLeuLeuProMetLeuGluMetLeuAlaGlnAlaTrpGlnSerTyrSerGln	82
DB	2811	GCCGACTATGCGTCTGTTGGCGCAATGCTGGCGGAGGTGGCGGCACATCGCCCGCAA	2870
QY	83	ArgAsnSerSerLeuLysAspIleAlaArgGluAsnAsnProAspTrpIleLeu	102
DB	2871	CGCCCGCAAGCCTTGCACAGCTAGACAGCAGCGG-----CACGCGCGCTGTTGGCG	2924
QY	103	SerAsnLysGlnValGlyGlyValCysTyrValAspLeuPheAlaGlyAspLeuLysGly	122
DB	2925	CAACGACACATGCTCGGCTACAGCGCTACGACAGCGCTTTCGCGCAGCTTGCAGGC	2984
QY	123	LeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeuMetPro	142
DB	2985	GTGCCCGCGCGCGCGCTATCTGCAGGAACCTGGCGGTGCTACCTGCTGCTGCTGCA	3044
QY	143	LeuPheLysCysProGluGlyLysSerAspGlyTyrAlaValSerSerTyrAsp	162
DB	3045	TTCTGCGCGCAGCGCGCGGACACGACGCGGCTTTGCGGTCAGCGATTACGCCAG	3104
QY	163	ValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaLeuHisGlu	182

```

Db 3105 GTGGAACCAAGCTGGCGACCAACGACCTGGTGGCGCTTACCGCGCGCTGGCGGAA 3164
Qy 183 AlaGlyIleSerAlaValValaspPheIlePheAsnHisThrSerAsnGluHisLys 202
Db 3165 GCGGATCAGCTGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3224
Qy 203 AlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePheProAsp 222
Db 3225 GCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3284
Qy 223 ArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGlnHisPro 242
Db 3285 CGACCCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3344
Qy 243 GlyGlyPheSerGlnLeuGluAsp---GlyArgTyrPheValTyrPheAsnSerPhe 261
Db 3345 GGCACCTTCACCTGGGTGGACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3404
Qy 262 GlnTyrAspLeuAsnTyrSerAsnProTyrPheValPheArgAlaMetAlaGlyGluMetLeu 281
Db 3405 CAATGGGATTTGAACCTGGAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3464
Qy 282 PheLeuAlaAsnGlyValAspIleLeuArgMetAspAlaValAlaPheIleTyrLys 301
Db 3465 CGGCTGGCAACCTGGCGGTGGAACATTCGCCCTGGATTCACCGCGCTATCTGTGAAA 3524
Qy 302 GlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPheAsn 321
Db 3525 CGCATCGGCACCGATTCATGATGAATCAACCGGAGCGCACACCGCTGCTGTGGCGTGG 3584
Qy 322 AlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHisPro 341
Db 3585 GCGGTAAACCGATATCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3644
Qy 342 AspGlnValValGlnTyrIleGlyGln-----AspGluCysGlnIleGly 356
Db 3645 ACGCAACTGCGCGCTACTTCGGCAGCGCGGTGGATAGGCGGCAAGATGTCATGCGCC 3704
Qy 357 TyrAsnProLeuGlnMetAlaLeuLeuTyrPheAsnThrLeuAlaThrArgGluValAsnLeu 376
Db 3705 TACCACAGCAGCGCTGATGCGCAGCGCGCTGCTGGCAGCTGGCGTTCACAGCGCGGACATC 3764
Qy 377 LeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHisThrAlaTyrValAsnTyr 396
Db 3765 CTGCACACGCTGATCGCGCACAGCCACCATTCGCGCGCCATGCTGCTGCTGCTGCTG 3824
Qy 397 ValArgSerHisAspIleGlyTyrThrPheAlaAspGluAspAlaAlaTyrLeuGly 416
Db 3825 GTGCGTTGCATGACGATATCGCTGGAATGCTGCTGCAGCAGCAGCGCTGCGGCAATGCC 3884
Qy 417 IleSerGlyTyrAspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGly 436
Db 3885 GGCACACCGCGCTTTCGCTCGCGATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3944
Qy 437 SerPheAlaArgGlyValProPheGlnTyrAsnProSerThrGlyAsp-----CysArg 454
Db 3945 AGTTACGACCGCGCGGAGATTTCCAG-----AGCAGCGCGCGCGCGCGCGCGCG 3995
Qy 455 ValSerGlyThrAlaAlaAlaLeuValGlyLeu-----AlaGlnAspAspProHis--- 471
Db 3996 ACCAAGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4055
Qy 472 -----AlaValAspArgIleLysLeuLeuTyrSerIleAlaLeuSerThr 486
Db 4056 GCGGCGTTGGCGGTTCAGTGCAGCGCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 4115
Qy 487 GlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThrLeuAsnAspAspTrp 506
Db 4116 CCGGCGGTTCGCTGATCTACATGGCGGACGAATTTGGGATGGTCAATGACCGCGCTAT 4175
Qy 507 SerGlnAspSerAsnLysSerAspAspSerArgTyrAlaHisArgProArgTyrAsnGlu 526

```

```

Db 4176 CGCCACGACCGCGCACCGCCAGCATGAAGCGCGTGGCTGTCATCGCGCGCGATGG 4235
Qy 527 AlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGlyGlnIleTyrGlnGlyLeu 546
Db 4236 CAGCTTCCGCGCGAGCTGACGATGCGACAGCTGAGTGCAGCGGTGTCAGCGCTTG 4295
Qy 547 ArgHisMetIleAlaValArgGlnSerAsnProArgPheAspGlyGlyArg---LeuVal 565
Db 4296 CGTGATGTATCCGACACGTCGCGCGTGGCGCGTGGCGAGCAGATCAACGACATGGCA 4355
Qy 566 ThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArgAsnAlaLeuLeuAla 585
Db 4356 AGCATGGCGCTGAATGATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4415
Qy 586 PheGlyAsnPheSerGluTyrProGlnThrVal-----ThrAla 598
Db 4416 TTGCATAACTTTTCAGCAGCAGCGCTGCGAGCTGCGAGCTGCGAGCATGGCATGGA 4475
Qy 599 HisThrLeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyLysThrValSer 618
Db 4476 TGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4523
Qy 619 LeuAsnGlnAspLeuThrLeuGlnProTyrGlnValMetTyrLeuGlu 634
Db 4524 GCGCATGATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4571

RESULT 12
LOCUS AE012454 11605 bp DNA linear BCT 23-MAY-2002
DEFINITION Xanthomonas campestris pv. campestris str. ATCC 33913, section 362
of 460 of the complete genome.
ACCESSION AE012454 AE008922
VERSION AE012454.1 GI:21114602
KEYWORDS
SOURCE Xanthomonas campestris pv. campestris str. ATCC 33913
ORGANISM Xanthomonas campestris pv. campestris str. ATCC 33913
Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
Xanthomonadaceae; Xanthomonas.
REFERENCE
1 (bases 1 to 11605)
da Silva A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,
Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida
Jr., N.F., Alves, I.M.C., do Amaral, A.M., Bertolini, M.C.,
Camargo, L.E.A., Camarotte, G., Cannavan, P., Cardoso, J.,
Chamargo, P., Chapina, L.P., Cicarelli, R.M.B., Coutinho, L.L.,
Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S.,
Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,
Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite
Jr., R.P., Lemos, S.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,
Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,
Mench, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,
Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira, J.R., H.A.,
Rossa, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F.,
Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos
Santos, M., Truffi, D., Tsai, S.M., White, P.F., Setubal, J.C. and
Kitajima, J.P.
Comparison of the genomes of two Xanthomonas pathogens with
differing host specificities
Nature 417 (6887), 459-463 (2002)
JOURNAL Nature 417 (6887), 459-463 (2002)
MEDLINE 22022145
PUBMED 12024217
REFERENCE
2 (bases 1 to 11605)
da Silva A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,
Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida
Jr., N.F., Alves, I.M.C., do Amaral, A.M., Bertolini, M.C.,
Camargo, L.E.A., Camarotte, G., Cannavan, P., Cardoso, J.,
Chamargo, P., Chapina, L.P., Cicarelli, R.M.B., Coutinho, L.L.,
Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S.,
Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,
Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite
Jr., R.P., Lemos, S.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,
Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,
Mench, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,
Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira, J.R., H.A.,

```

Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinoia, L.A.F., Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos Santos, M., Truffi, D., Tsai, S.M., White, P.F., Setubal, J.C. and Kitajima, J.P.

# Direct Submission

Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de São Paulo, Av. Prof. Lineu Prestes 748, São Paulo, SP 05508-900, Brazil

## FEATURES

### Location/Qualifiers

1. .11605  
/organism="Xanthomonas campestris pv. campestris str. ATCC 33913"  
/mol\_type="genomic DNA"  
/strain="ATCC 33913"  
/db\_xref="ATCC:33913"  
/db\_xref="taxon:190485"  
/note="pathovar: campestris"  
164. .1357  
/gene="oprO"  
/note="synonym: XCC3353"  
164. .1357  
/gene="oprO"  
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"  
/codon\_start=1  
/transl\_table=11  
/product="porin O"  
/protein\_id="AA042623.1"  
/db\_xref="GI:21114603"  
/translation="MINDTLRRAITACAAICLAPAAHAADGEPVTAEGVRKWDPTFVNDERGTPERTQFTRVLDVAGFYGNFYAEAFAGLOVEPGSGILARDVYIAKTSAGTLTGQFQKSLDDRTGSNYGPELBERGYASTTLAPYKRAISWQTSRPD AWTSAAYSLGISDNTAGSGALGRLTLPAPMGDARLHLGVSLAHRYSHPGSGGA SLLIRPPENDLANNRSLTLPAPADGRDTEDEKSLVAVVWGPSPKALRYESGGLD DGAQNAKULASGLVSWFTGESRVDKTRFSRIKTPNRPSPGAFELALDYNRGD QHLDGQPFINASTUSWTLGGNWPYKPNLRVNLVINDSRNDRRTMNAVVDHTLAVTGR FQYD". .2910  
/gene="citM"  
/note="synonym: XCC3354"  
1582. .2910  
/gene="citM"  
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"  
/codon\_start=1  
/transl\_table=11  
/product="Vg++ citrate complex transporter"  
/protein\_id="AA042624.1"  
/db\_xref="GI:21114604"

### gene

1582. .2910

/gene="citM"

/note="synonym: XCC3354"

### CDS

/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"

/codon\_start=1

/transl\_table=11

/product="Vg++ citrate complex transporter"

/protein\_id="AA042624.1"

/db\_xref="GI:21114604"

/translation="MLTALGFGWITFMVIMSKSLPLVALITVPIVALLGPGTGNEMLDGKIKTAPGWMFLFYGVIMIDAGLFDPLVGRILRVKGDPLKIVMGTA ILAMLSLDGSGTMTVTSAMFLYQRLGNALNLTCTVTLASGVNMLTPWGQPTA RAATLHVPADVPFVLPAMAMAGIILALAWLGMERRRRLGVRLPDPGNMLETSLPDDNALPRVETEDMKRPKLLVNLVLTALMAALVVGILPMPVLMIGFALAMI NPYLAQRRLVNHAGVNSLVVSLIPAGTIGLSNTGMVEAMSRFLAVIPDSWG PHLAVITLVSPPFFSNDRAPFYGVLPILSRAAGHYGTPVEMARSLAQGPVHLL SPFVPSVLVLGKLVDFADHQRTFLKWAVALISMLAGGLLGLFLPLAR"  
3073. .3813  
/gene="phbB"  
/note="synonym: XCC3355"  
3073. .3813  
/gene="phbB"  
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"

### gene

3073. .3813

/gene="phbB"

/note="synonym: XCC3355"

### CDS

/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"

/codon\_start=1

/transl\_table=11

/product="acetoacetyl-coA reductase"

/protein\_id="AA042625.1"

/db\_xref="GI:21114605"

/translation="MTLRIAYTSGVSGVTGAI COKLARTGTTVAGCGPNSPKSAW LREQRELGFDFVASSGNADMSTWAAFAKVAEVEIDVLVNNAGSGRDLTFQMSR DDMNAIVASNLFSNTKQWVDGNTSRGMRIYVNIQSVSAKQIGQINPATAKAM HGFSRALAQEASRGVTVNTISPGVIASASISNFPDPVLDRLATVSPVVRIGKPAEVA

### gene

4181. .5212

/gene="cebr"

/note="synonym: XCC3356"

### CDS

4181. .5212  
/gene="cebr"  
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"

/codon\_start=1

/transl\_table=11

/product="transcriptional regulator"

/protein\_id="AA042626.1"

/db\_xref="GI:21114606"

/translation="MAKHPKATPAAPQWADIARMAGVSTSVRALAGSPVVAERT RAYIKQAAAGYQVDFVARSRAKRSNTVSVAVMMHSDQPSLDPMTMLLALAE ELTGRYSMLSKLNHRQDWQEQARSGSDGVILGQSEHAALDAADGJPMAY WGSRIDQSVISVSGDNFQGGALATEHLIASRRRIAPLGGDQLPEVAPRFAGYRQAL EOHGLEFTRLHARSHFLSDAYRLTRAMLKKTDPDGLFAASDVIAVGATRAIVEAG HRVPQDLSLNGFDDIPLAAYSQPLTTVQDGLGAARLLVDRLLALTAGEAVDSEMP VKLVVRESA"

complement(5440. .6759)

/gene="sucI"

/note="synonym: XCC3357"

complement(5440. .6759)

/gene="sucI"

/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"

/codon\_start=1

/transl\_table=11

/product="sugar transporter"

/protein\_id="AA042627.1"

/db\_xref="GI:21114607"

/translation="MSSTAPPLSPARLALNAGPFGVQYSPGLQOSNMSPITVNLGAD HAPLPTMLAGPITGLVLOPVGALSDRSVTRGCRMPYVVLGALVSLCLLAMPST ALMNAVCLWLDRAANNVNEPRALNSVDLAPQRPGLYLTQSAFSLCLLAMPST PLLVWQSDAANAHPVITIAAFVIGAGFSAASILLTARSVPFVPAAREARNR KAGTGLCATREBGSALRDMPTMRQAPVNLFWYALFISYQWVILSLSTLFGTTE ANSHGFAGLVNQIQGFYNTIAFLAAFAVPPVRRVGPKYTHAACLLAAGVGMVL PGIENRWLLLMIGIGLAWASMGPNYLMADSIPTPTGVYMGLEFLVLPMLIQ IVTLPVYPELLHGDPRNVIRLAGALMLAAVAMLCVRIRKPKAAVA"

7007. .9481

/gene="fyuA"

/note="synonym: XCC3358"

7007. .9481

/gene="fyuA"

/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"

/codon\_start=1

/transl\_table=11

/product="iron-dependent receptor"

/protein\_id="AA042628.1"

/db\_xref="GI:21114608"

/translation="MRERIAVSTLHRLHALACAVTTCCLAPLAAQDATPPAPATP PAADSAVNLDSVFTGTSTATTKLSSVSVTGAEAIQSAPRSTAEIERNIPGIR SESSGSGENAIARVGIPIVAGGAKFLQEDGLPVMFGLIAPGNADI ELRSDFTMD RAIKRGSGASTTSNAPGIIINFIKGTGTEGSGVSGRLDYDNRIDRINYGAPYA EHWQFNIGFPFRGDVDRGYTIDKGGQKAMLTRLFENGYVAVYKILNDRAGLIS PVTSVRGDRGSDLGGPFGDFDNDTLYSRNFRTPDVLGNWQPRRTDLDGMHPIS RTTGAZAFDLGNGWLSDFRIADNSGRVSPFPPEVTDAAALSSIGGAGAQVFA GGANAGQATGTATRLFNVAINDLGNVTNDSLSEFGDGKRTLRLMGYYTSRQD IDMDWTNYSVQSLGRDRLLNVVDAAGVSRSONGLVYGTGPFMGDCCTRSYDVRVD VNAVYSLTFDSKLSIDGSLRVDMGDARGNVSGTAIAQNLDVNGDGVIOPEQORVAT VDTANRPPVDYDNYLSLGGNYLINDDLGAPARVSRGARANADLLPGRVIRDGVS TSDEAVNVVRQTAGLKKRRDGLSLFATAFTQEQNFVTSORFNRSYKANGIEL EASYRYEGFTVNGGVWTDIARDQITPENTGNVPRQADFVWQLTPSRVGDGYQFG VNLITGTEAYTQDSNOLKMPGYQVNLFGDYRITDALTVALNVNLENFTTGLTBAEEA TIPANGIIRARSITAGRTTSLSLRYDF"

9515. .11428

/gene="XCC3359"

9515. .11428

/gene="XCC3359"

/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"

```
/codon start=1
/transl_table=11
/product="amylase or alpha amylase"
/protein_id="AAM42629.1"
/db_xref="GI:21114609"
/translation="MIASSPIDAALRASVAALDPSTAVATLARFDTHAPLLDALSL
TYLGHADVTVLWLPWLSALGVVAQAPALRLDSSRAPRFGQDMGLGSAFYDRP
ACTLRGVAERVPYLOELGVYRLHLEPFLPARAGDNDGGFANSDYQVPEFALGSDNLIY
ALTARLRANISLCAFDVLIHNTADDAWAQARAGTRLYDYHFFAURNAPDQYDTT
```

## Alignment Scores:

```
Pred. No.: 2,3e-86 Length: 11605
Score: 1116.50 Matches: 241
Percent Similarity: 56.51% Conservative: 106
Best Local Similarity: 39.25% Mismatches: 236
Query Match: 32.76% Indels: 31
DB: 1 Gaps: 10

US-09-843-007A-2 (1-636) x AE012454 (1-11605)

QY 43 ArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGlyAsnAsn 62
DB 9605 CGTTTCGATACGATGCCCGGCTTCGACGGCTGTCCACGCTCTACGGCGACCAT 9664
QY 63 GluAlaLeuLeuProMetLeuLeuMetLeuLeuAlaGlnAlaTrpGlnSerTyrSerGln 82
DB 9665 GCCGACTACGTCACCTGCTGCACAGTGGCTGAGCGCACTGGCGTGTGCACAGGCA 9724
QY 83 ArgAsnSerSerLeuLysAspIleAspIleAlaArgGluAsnAsnProAspTrpIleLeu 102
DB 9725 CGGCACCGGCTTCGGGCACTTCGACGACACCGCGCA-----CCGGGCTGTTCGGC 9778
QY 103 SerAsnLysGlnValGlyValCysTyrValAspLeuPheAlaGlyAspLeuLysGly 122
DB 9779 CAGCAGGACATGCTGGCTACAGCGCTACGTGGACCGCTTTTCCGGCAGCTGGCGGT 9838
QY 123 LeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeuMetPro 142
DB 9839 GTTCCGACGCGGTTCCGATCTGACGGAACCTGGGCGTGGGTATCTGCACCTGTGCA 9898
QY 143 LeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyrArgAsp 162
DB 9899 TTCCTGGCGCAGCGCGCGCACACGACGCGGTTTGGTTCGCTCAGCATACGGCCAG 9958
QY 163 ValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAlaLeuHisGlu 182
DB 9959 GTGGAACCGGCTTCGGCAGCAACGACCATCTGTGCACTCACCACCGCGCAGCTACGCG 10018
QY 183 AlaGlyIleSerAlaValAlaValAspPheIlePheAsnHisThrSerAsnGluHisLeuTrp 202
DB 10019 GCCAACATCAGCTGTGGCAGATTCGTGTCTCAACCCACACCGCGCAGCACCCGCTGG 10078
QY 203 AlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePheProAsp 222
DB 10079 GCACAGCGCGCAGCGCTGGCGATACGGCTTATCTCGACTACTACCATCATTCTGCCGAC 10138
QY 223 ArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGlnHisPro 242
DB 10139 CGCAACGACCTTCGACGATACACACACCCCTGGTGCAGGTGTTCCCGCAGACTGCGGCC 10198
QY 243 GlyGlyPheSerGlnLeuGluAspGlyArg---TrpValTrpThrPheAsnSerPhe 261
DB 10199 GGCACATTCACCTGGTGGCAGCAGACCGCGGAGTGGATGGACACCGCTTATCCCTAT 10258
QY 262 GlnTrpAspLeuAsnTyrSerAsnProTrpValPheArgAlaMetAlaGlyClnMetLeu 281
DB 10259 CAGTGGGATTTGAACCTGAGCAACCGCGGTGTTTCGGCGAGATGGCATTCGGCATGCTG 10318
QY 282 PheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTrpLys 301
DB 10319 GAGCTGGCAATCTGGGGTGGAGCGGTTCGACTCGATTCCAGCGCGTACTCTGGGAAA 10378
QY 302 GlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPheAsn 321
```

RESULT 13  
AE005791/c

LOCUS  
DEFINITION  
10822 bp DNA linear BCT 12-JUN-2002  
Caulobacter crescentus CB15 section 117 of 359 of the complete genome.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AE005791 AE005673  
AE005791.1 GI:13422441

REFERENCE  
AUTHORS  
Nierman, W.C., Feldblyum, T.V., Leub, M.T., Paulsen, I.T., Nelson, K.E., Eissen, J., Heideberg, J.F., Alley, M.R., Ohta, N., Maddock, J.R., Potocka, I., Nelson, W.C., Newton, A., Stephens, C., Phadke, N.D., Ely, B., DeBoy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L., Haft, D.H., Kolonay, J.F., Smit, J., Craven, M.B., Khouri, H., Shetty, J., Berry, K., Utterback, T., Tran, K., Wolf, A., Vamathevan, J., Ermolaeva, M., White, O., Salzberg, S.L., Venter, J.C., Shapiro, L. and Fraser, C.M.  
Complete genome sequence of *Caulobacter crescentus*  
Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)  
21173698  
11259647

REFERENCE  
AUTHORS  
Nierman, W.C., Feldblyum, T.V., Paulsen, I.T., Nelson, K.E., Eissen, J., Heideberg, J.F., Alley, M.R., Ohta, N., Maddock, J.R., Potocka, I., Nelson, W.C., Newton, A., Stephens, C., Phadke, N.D., Ely, B., Leub, M.T., DeBoy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L., Haft, D.H., Kolonay, J.F., Smit, J., Craven, M., Khouri, H., Shetty, J., Berry, K., Utterback, T., Tran, K., Wolf, A., Vamathevan, J., Ermolaeva, M., White, O., Salzberg, S.L., Shapiro, L., Venter, J.C. and Fraser, C.M.  
Direct Submission  
Submitted (31-Jan-2001) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA  
Location/Qualifiers  
1. 10822  
/organism="Caulobacter crescentus CB15"  
/mol\_type="genomic DNA"  
/strain="CB15"  
/db\_xref="taxon:190650"  
141..452  
/gene="CC1128"  
141..452  
/gene="CC1128"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK23112.1"  
/db\_xref="GI:13422442"  
translation="MKRITIAAGVAGSPFIAQALAGGSPEDQWACIQCHNTRDA  
AQMLECNADGGKLSKPCVVDNSAAGKGFDAKMCIAEKLPGAKTGTVKVPFRPGG  
A"  
670..1176  
/gene="CC1129"  
670..1176  
/gene="CC1129"  
/note="identified by match to protein family HMW"  
/codon\_start=1  
/transl\_table=11  
/product="RNA polymerase sigma-70 factor, ECF subfamily"  
/protein\_id="AAK23113.1"  
/db\_xref="GI:13422443"  
translation="MSDLASLRQRDLQYVRRTRDPATAEDLVQETFERLLAYGQ  
AKTVPRALGVGIALNLVDRHFLAQRRAFDLDDIPCDDPAPQILMHRQKVEVF  
SOVLDAWPLLRDVFRRRLHGHSSRQISKDLSLSEAAVEKHVARALEQLREVARAE  
RAGGGRS"  
1173..2096  
/gene="CC1130"  
1173..2096  
/gene="CC1130"  
/note="similar to SP:P23485 GB:M63115 PID:145931

PID:537133 GB:U00096; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="transcriptional regulator, putative"  
/protein\_id="AAK23114.1"  
/db\_xref="GI:13422444"  
translation="MNRIOAEALWAARRLDADAQDPDAFPAWKAADPANARAYEQVW  
RASDDPALSEALRYSEORFAAPRAAPVLAGGLACGLAVFLAWPAQVMTVOPOV  
IQTAGQQRVALADCTQVTLDGATRLDVLGRQVRLYRGENAFVDAHARPEFI  
VQAEGRSVLTGADFLERGGKRLAPAGTGVIRRTAELTVGQAFPAKEG  
RLSVRTDPDDQDWRGTGLETGVTGLARLVRLNRASDKPTIIDPALGRQVARGF  
RLDDEALARNALMHGFTIETSENGLELR"  
2172..5102  
/gene="CC1131"  
2172..5102  
/gene="CC1131"  
/note="similar to GP:9187833; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="TonB-dependent receptor"  
/protein\_id="AAK23115.1"  
/db\_xref="GI:13422445"  
translation="WTFPRDGDNDTAKTWTWLGASASLAKMLACAPALAPAAKPTIL  
AIPAGDATALPYFSRATGLQVLADPALKKGRTEGVGRQFEARGLTELRGTGLNF  
VRNGAAITLRPAQQRVSTAPAAQAAPVATAQAQAEPAWLEALVWVGFPANSLAR  
ALADKRAANVVDIAISADIKGFPNNIAEALQVPGVATIDRGEGLFVRVGLGPN  
FQVITLGRSAANVMNRDSCQSGRQFDTLPSELVAGVGVKSPGLASDEGALGV  
DDVETRLDGRSTLALSAATSRPOLAGTKDPRLSGLASWNTAEGTGVLAALVPDQ  
RTLQDRITGVGWSYADPASTLTKMTIAQGISLYPTAIRPTLEREDRERKGVLSAQY  
RPSDETHIVDASVTKLDHYDELTVSDQNVAILAGSAVTKQGLVGGTVNSTQIT  
GREVDLADNLVAGRIQKLEDWTLADIAARAFSETNKPIRRLGLPGITTF  
SMTVDDDAVPSVRLATADLGNPSLPPRIEMRENNATDDETLQFPAERPLSGLL  
SRVQGVKRYERSDYNRADINTYTNLAGKFFGAFYDAFVSFLGVDGTLPNNWA  
MPDAFWGLADKTQRATARDKNSYQVSDIAAGYILGDVSEKVFNPALRGQLGVR  
VARTEQISAGHADNGSAIPVRYKSTYTVLPTANFALELTDLQARFAAKVITRPS  
LADLAPRLTNSGTVLVANGNPILKPFPAQVDTLEWYFAGSLIAGAFYKDIIT  
TFVYSQKTNIVDQTVLITLSTPTNGSKAKIGIELAQHLFKRLPAPFDGLGFLANYT  
ITDEATYSPTLKDMVNVAKNSPLTAFYEKDRFAAKASVSWDDVDLQDVGKGLSA  
LNDSPAGSLDASVSKITPSTIAAQAQNLNLSAQWOPVKGWFGFYTHYGRITLSFGL  
RAKF"  
5208..6737  
/gene="CC1132"  
5208..6737  
/gene="CC1132"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAK23116.1"  
/db\_xref="GI:13422446"  
translation="WLKLSHRRPCLSPVALVALCLAPVSLPAFAHGHAVQVGP  
SRPAGRAVAKSERKWAQDHHVSHFVSLKPGADPSQPPALIKAGDAVNEITNARM  
ALKHGLAAWATDHGPNHNSKLNDEAPPELLRSRVETPAVLQWGMELDTPAHSHS  
LILPSEERDALGIESRNVNREPAPADPAQDEAFWLEALRKALKKPEVWIANHS  
PSRGATAVRGYGLVSPSELNNNDTADPAIMEGAPGHQASLKPLGALPKQGDGGG  
YSRAPTMGGFDQMTAKLGFMDMLGEGRRWITATSDSHRHYTDGGDFWPGYSKT  
YVKAHTYDDVLGDLGRVFTVGLVSEVETASVAGAKADIGRLEVAQSDVTY  
VIRVDPAGKNFGSPSVARVDLIQGLGPATDRSDTANPTTRVIRPAFTPSHARE  
GEMLSMTHLVKDVRRPVYIRVRCNGAEAPLPDGEDPWSDLFWFANPFIITVASA  
R"  
complement (6759..7994)  
/gene="CC1133"  
complement (6759..7994)  
/gene="CC1133"  
/note="identified by match to protein family HMW"  
/codon\_start=1  
/transl\_table=11  
/product="transporter, putative"  
/protein\_id="AAK23117.1"  
/db\_xref="GI:13422447"  
translation="WTQTVSDASDRPLRILITVLMFMFPAMTITDSVGVILPHVIKTF







```

DREGEMARELVCHRYGVEQVQLWLSALDDASIRKALRLPLFGHETFNLYHSALGRS
RADWLGMMNRULFTLLRGSGYGVGLPVGRVOTPTLRVLVDVDRDSIANFVVPKAI
DVOLEHAGORFNAWRAPEDACDQGRCLQSLAQOARADIHGAGTARAVKATERYR
EAPLPFDLGTLOELCSKFKGLGAQETLIDIAQLYTHKLIITYPRSDCCGTLFQSQAE
APAILARADASLALDLOPYLEPQRSFARNDAKVAHGHILPTTAASDPPLPAKH
KAVYTLIRARYLAQFPNEHYDKTQADFCAGHALRAVGKQIVFPGWRALPEALTPA
KGEAPPAQVLPALRGQCNVHGLQDLQTLWTPPKPTFGTEGLIRAMKNVAKLIDDP
LQKQKETTIGTETRASTIIOGLDLGVLRNGKALSATPAFSLIDVPRADDPG
TTAIFQALDMQSGEMPLEEFVARQSAWMLKVERCGRMTISGPAAGAAPPWKKK
RRGGGSKATGSKATPKRQARKKRAPI"
5414..6703
/locus_tag="PP4020"
5414..6703
/locus_tag="PP4020"
/notes="similar to GP:9951625; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="oxidoreductase, putative"
/protein_id="AA069614.1"
/db_xref="GI:24985720"
/translation="MFKQAEHVASVYAGTYTNPRLPTLQGTQDADVLIIAGFSG
LATALRLTQAGRVTLLEASRVAAAGSGGQALLGSCMDPPELQALGIERTLRW
ASMWAAEMRRLPQRHGFIDYRQGSILWTAVLPRVKMLBEALHDAHKGYDALRL
IGRDELPOINDSPRYQALYDAKGAHLNPLKLAQGLASTIEAGQIYEQSOVLVYQQ
ARDGYARDRGEVRCQVLVLACNATIDRLDGLSRLLPVGSYQVATAPLDADFAHS
LUPRNSVDINDQFVDPYFLKTLTDHRLLLFGGCTYLGGLPGKIPDVASATRYLERYFPOLA
GVADYAMGHHIDCSIKRTPDVREGORYLWQFSGHGVLPPTLAARAVSDAILGDED
LLAYQGDINDGFPFGCDLLAALPAAAKAWYMRDRV"
6876..7706
/genes="est"
/locus_tag="PP4021"
6876..7706
/genes="est"
/locus_tag="PP4021"
/notes="similar to GB:U07989, SP:P01596, SP:P01602,
SP:P01604, PID:1335148, PID:1619960,
PID:185945, PID:185947, PID:185949, PID:468243,
PID:470602, PID:470608, PID:470612, PID:561696,
PID:561705, PID:563650, PID:619427, PID:619601,
PID:619602, PID:619613, PID:619615, PID:619616,
PID:632988, PID:642582, PID:663000, PID:663001,
PID:722576, PID:758085, PID:762937, PID:790811,
PID:860994, PID:896278, GB:U07989, SP:P01596, SP:P01602,
SP:P01604, PID:1335148, PID:1619960,
PID:185945, PID:185947, PID:185949, PID:468243,
PID:470602, PID:470608, PID:470612, PID:561696,
PID:561705, PID:563650, PID:619427, PID:619601,
PID:619602, PID:619613, PID:619615, PID:619616,
PID:632988, PID:642582, PID:663000, PID:663001,
PID:722576, PID:758085, PID:762937, PID:790811,
PID:860994, and PID:896278; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="esterase"
/protein_id="AA069615.1"
/db_xref="GI:24985721"
/translation="MSVYTTKDGVOIEYKDKGPRDAPVTHHHGKPLSADWDQMLF
FLAEGYRVYADRHGCHGSSQVWNGHDMHTADDOVAUVSLGTQGVAVHVGSTGGGE
VRYWARYPEBKVAKXAVILAPVPLWQVPGGLPKSVFDFQQAQVSNRQAFYDP
VFTGYFGRNPGVDASGIIIGNWRQMGISAKAHYDGIYAFSQDFTDLKGIQOP
VLVHGGDDQIVPYNSGVLSAKLLPNTGLATYEGYPHGMPHTHADVINADLLAIQIS
"
7826..8617
/locus_tag="PP4022"
7826..8617
/locus_tag="PP4022"
/locus_tag="PP4022"
/codon_start=1
/transl_table=11
/product="conserved domain protein"
/protein_id="AA069616.1"
/db_xref="GI:24985722"

```

```

/translation="MLIRAKYDEITQILNLVLTMLTKQTQPDPSAASLTNALRHFAYS
CIGYXPQFLLATFNQYIDLFSGHYQLGSYRTNTGLMRPHSPDSPPDGGLN
AYACSGDFINVDPSGHGVLKSKSLPKTKIMORREALKAINENPVLBAFQGIA
ENSVTKKSLTSDTKALNQDILMKNKLIISIDKQGLKYKLMDDFSLRQIILED
VMKNDKTLNLSLSAYPRANSPEPSLSSASLSPATNKLTTRTKGN"
8693..9337
/locus_tag="PP4023"
/notes="This region contains an authentic frame shift and
is not the result of a sequencing artifact; this gene is
interrupted by an IS element.; membrane protein, putative,
interruption-N; identified by match to PFAM protein family
HMM PF01925"
9347..11387
/insertion_seq="IS0024"
9432..9791
/locus_tag="PP4024"
9432..9791
Alignment Scores:
Pred. No.: 4,8e-35 Length: 304517
Score: 550.00 Matches: 164
Percent Similarity: 43.72% Conservative: 83
Best Local Similarity: 29.03% Mismatches: 228
Query Match: 16.14% Indels: 90
DB: 1 Gaps: 20
US-09-843-007a-2 (1-636) x AB016789 (1-304517)
QY 118 GlyAspLeuLysGlyLeuLysAspLysLeuProTyrPheGlnGluLeuGlyLeuThrTyr 137
DB 56457 GCGCATTTCCGCGCGCTGATCATCAGCAAGCTGACTATATCGCGGAGTGGTGGTCAACACC 56398
QY 138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaVal 157
DB 56397 TTGTGGCTGTCGCGCTCTAT-----CCCTCGCCACCGCTGAGCGGTACGACATC 56344
QY 158 SerSerTyrArgAspValAsnProAlaLeuGlyThrLeuGlyAspLeuArgGluValIle 177
DB 56343 GCCGATATCAAGGCGCTGACCTGACTACGCGAGCATGCGCGGCGCGCGCGCTTTCATC 56284
QY 178 AlaAlaLeuHisGluAlaGlyLysSerAlaValValAspPheIlePheAsnHisThrSer 197
DB 56283 GCCGAGGCCACAGCGCGGCTCGGGTGTATCCGAGTGTGTTCATCACCACACACCGAGC 56224
QY 198 AsnGluHisGluTrpAlaGlnArg-----CysAlaAlaGlyAspProLeuPheAsp 214
DB 56223 GACCAGCACCGCTGTTCCAGCGCGCGCGCGCCAGCCAGCGCGCGCGCGCGAG 56164
QY 215 AsnPheTyrTrpIlePheProAspArgArgMetProAspGln---TyrAspArgThrLeu 233
DB 56163 -----TTCACGTTGTGTGCGAC-----GACGACCAAGAGTACGACGCGACG--- 56122
QY 234 ArgGluIlePheProAspGlnHisProGlyGlyPheSer---GlnLeuGluAspGlyArg 252
DB 56121 CGCATCATCTTCCTCGACACCGAGAGTCCCACTGGGACCTGGGACCGCTGCGCGGCGAG 56062
QY 253 TrpValTrpThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsnProTrpVal 272
DB 56061 TACTTCTGGCACCGGTTCTACTCCACGACGCGGACCTCACTTCGACAAACCGCGAGTG 56002
QY 273 PheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeuArg 292
DB 56001 CTCGAGCGGCGTGATCGCGCTCATCGGCTTCTGGCTGAGCTGGTGTAGATGGCTTCGGG 55942
QY 293 MetAspAlaValAlaPheIleTrpLysGlnMetGlyThrSerCysGluAsnLeuProGln 312
DB 55941 CTGACGCGGATCCCTTACCTGTGATCGAGCGGCGGACCAACCAACGAAAAACCTCGCGAA 55882
QY 313 AlaHisAlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaPheAlaValPhe 332
DB 55881 ACCCAGCATGTCTCAGGCGATCCGCGCGAGATCGACCGCACTACCCCGCATCGCATG 55822

```



```

QY 333 PheLysSerGluAlaIleValHisProAspGlnValGlnTyrIleGlyGln----- 350
DB 55821 CTGCTGGCGGCGAAGCAACCGAGTGGCTGAAGACACCGCCCGCTACTTCCTGGCGAAGCGGCAC 55762
QY 351 ---AspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTyrAsnThrLeu 369
DB 55761 GCGGAGCATGCCATGCTTCCACITTCCTGGTGGATCGCGCATGTACATGCGCTTG 55702
QY 370 AlaThrArgGluValAsnLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGlu 389
DB 55701 GGCATGGAAGACCGTCTCCCTATCCGACATCCCTGCGCAGACCCCGGAAATCCCGGCC 55642
QY 390 HisThrAlaTyrValAsnTyrValArgSerHisAspIleGlyTyrThrPhe---Ala 408
DB 55641 AACTGCCAGTGGCGATCTTTTGGCAACCATGATGAGTACCTGGAATGCTCAC 55582
QY 409 AspGluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsnArg 428
DB 55581 GACCGGAGCGGACTACCTG----- 55561
QY 429 PhePheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsnPro 448
DB 55560 -----TGGAACTACTACGCC 55546
QY 449 SerThrGlyAspCysArgValSer---GlyThrAlaAlaAlaLeuValGlyLeuAlaGln 467
DB 55545 GAAGACCGCGCGCGCATCAACTGCGCATTCGCGCGCGCTGCGGCCCTGCTGCAG 55486
QY 468 AspAspProHisAlaValAspArgIleLeuLeuTyrSerIleAlaLeuSerThrGly 487
DB 55485 COTGACCGTCGG-----CGCATAGAGCTGCTGACCGAGCTGCTGTGTCGATGCC 55435
QY 488 GlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr-----LeuAsn 502
DB 55434 GCACGCGCAACCTGTATTACGGCGACGATAGCATGGCGGACAAATATCTATCTGGT 55375
QY 503 AspAspAsp-----TrpSerGlnAspSerAsn----- 511
DB 55374 GACCGCGATGGGCTTGGCACACCCATCGAGTGGTGGCGACACCGCAACGCGCGTTCCT 55315
QY 512 LysSerAspAspSerArgTrpAlaHisArgProArgTyrAsnGluAlaLeuTyr----- 529
DB 55314 CCGCGCGACCGCGCGCTGGTG---CTGCGCGCGATCATGACCGCGCTCTACGGTTAC 55258
QY 530 -----AlaGlnArgAsnAspProSerThrAlaAlaGlyGlnIleTyr 543
DB 55257 CAGACTGTCAACGTGCGAGCGCAAGCCACGACCGCATTC-----CTGCTG 55210
QY 544 GlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAspGlyArg 563
DB 55209 AACTGACCGCGCATGCTGGGTGCGGACGACGACGACGACGACGACGACGACGACGACG 55150
QY 564 LeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArg----- 579
DB 55149 CTGCGTACCTCAGCGCCAGCACCGCCGATCTTGCCTATATCCGCGATATACCGAC 55090
QY 580 -----AsnAsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThr 595
DB 55089 GCGGAGGCAATACCGAGTGTATCTCTGCTGGCGCAAGTGTCTGCGCGCCCGCCAGGCT 55030
QY 596 ValThrAlaHisThrLeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyLys 615
DB 55029 GCGGAACTCGAATGTGCATATACCGCGACAAAGTACCGGTGAGATGCTGGGTGCGAC 54970
QY 616 ThrVal-----SerLeuAsnGlnAspLeuThrLeuGlnProTyrGlnValMet 631
DB 54969 GCGTTCCCGCCAAATAGCGCAACTGCGGTTCTGCTGACCTTCCACCTTATGCTGTTCTAC 54910
QY 632 TrpLeuGluIleAla 636
DB 54909 TGGTTCTCTGCTGCC 54855

```

RESULT 15

```

AL646077/c
LOCUS
DEFINITION
  Ralstonia solanacearum GMI1000 megaplasmid, complete sequence;
  segment 2/11.
ACCESSION
  AL646077 AL646053
VERSION
  AL646077.1 GI:17430642
KEYWORDS
  SOURCE
  ORGANISM
  Ralstonia solanacearum
  Ralstonia solanacearum
  Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
  Burkholderiaceae; Ralstonia.
REFERENCE
  1
  AUTHORS
    Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
    Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
    Chandler,M., Choisme,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
    Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schiex,T.,
    Sguier,P., Thebaud,P., Whalen,M., Wincker,P., Levy,M.,
    Weissenbach,J. and Boucher,C.A.
  TITLE
    Genome sequence of the plant pathogen Ralstonia solanacearum
  JOURNAL
    Nature 415 (5871), 497-502 (2002)
  MEDLINE
    21681879
  PUBMED
    11823852
  REFERENCE
    2 (bases 1 to 189050)
  AUTHORS
    Boucher,C.A.
  TITLE
    Direct Submission
  JOURNAL
    Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
    Crémieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
    Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
    BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
    Dausset-CEPH, 27 rue Juliette Dedu, 75010 Paris, France, IMCM CNRS
    URGV, 2 rue Gaston Crémieux, CP5706, 91057 Evry Cedex, France,
    Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
    F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire
    INRA, BP27, F31326 Castanet-Tolosan Cedex
    Christian.Boucher@toulouse.inra.fr
  COMMENT
    http://sequence.toulouse.inra.fr/R.solanacearum.html.
  FEATURES
    Location/Qualifiers
      1..189050
        /organism="Ralstonia solanacearum"
        /mol_type="genomic DNA"
        /strain="GMI1000"
        /db_xref="taxon:305"
        /plasmid="megaplasmid"
        repeat_region
          79..1500
            /note="RS49+ or RS06146
            LONG DIRECT REPEAT 49 +
            predicted by Homology"
            /evidence=not_experimental
          83..2971
            /gene="RSP0175"
            /note="synonym: RS04687"
            83..2971
            /gene="RSP0175"
            /functions="miscellaneous; hypothetical/global homology"
            /note="product confidence : putative
            Gene name confidence : hypothetical
            predicted by Codon_usage
            predicted by Homology
            predicted by Framed"
            /codon_start=1
            /evidence=not_experimental
            /product="PUTATIVE VGR-RELATED PROTEIN"
            /transl_table=11
            /protein_id="CAD17326.1"
            /db_xref="GI:17430643"
            /db_xref="GOA:Q8XID8"
            /db_xref="SPTREMBL:Q8XID8"
            /translation="MDFKTLIQHLPAPHLRYALEGEGPTAEALVAEAWLGRALSELFL
            EWRVAASANARFALESFIGQRTVLTTLADGTQARTGLRQAEQAGDSGLARYRL
            TVPWFVLAQCRHSQVFQNRPLADILEVLSPYAVAAWRFPAAGADRMFAAFGRSH
            IAFRETQYRFTVRLAELAGLTTFVEQAPSGHALLIFADSPRIEDTASRAEGGI
            RYRRAHSGERDAIQALICHSRTAVGGVAAWDAQAQKRAFHVPFRFCGIAGSPDA

```



Qy	158	SerSerTyrArgAspValAsnProAlaLeuLeuGlyThrIleGlyAspLeuArgLeuValIle	177
Db	92764	GCCGAATAACCGCGCGCTGCACCCCGACTACGCGACCATGGCGCGACGCGCGCTTCATT	92705
Qy	178	AlaAlaLeuHiGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSer	197
Db	92704	GCCGAGGCGACCGCGCGCTCTGCCGTCATTACGAGCTGGTCATCAACACACACTCG	92645
Qy	198	AsnGluHisGluTrp-----AlaGlnArgCysAlaAlaGlyAspProLeuPheAsp	214
Db	92644	GACCAGACCCGTGTTTCAGCGCGCGCGCGCGCAAGCGCGGCTCCGCGCTGCGCGAC	92585
Qy	215	AsnPheTyrTyrIlePheProAspArgArgMetProAspGlnTyrAspArgThrLeuArg	234
Db	92584	-----TTCTACGTGTGGTCGGACCCAC-----GACAGAAATACGCGCGACCGCGC	92540
Qy	235	GluIlePheProAspGlnHisProGlyGlyPheSer-----GlnLeuGluAspGlyArg	252
Db	92539	ATCATCTTCATGCACACCGAGCCCTCCAACTGGACTGGACCCGCTGGCGCGACGCC---	92483
Qy	253	TrpValTrpThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsnProTrpVal	272
Db	92482	TACTACTGGACCGCTTCTATTTCGACACGCGCGACCTGACTGATTCGACAAACCCGCGGTG	92423
Qy	273	PheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeuArg	292
Db	92422	CTCAAGCGCGTGTGGCGTGTGAAGTTCTGGCTCAACCTGGCGGTGGACGGCTCGCG	92363
Qy	293	MetAspAlaValAlaPheIleTrpIleTrpIleGlnMetGlyThrSerCysGluAsnLeuProGln	312
Db	92362	CTCGATCGCGTGCCCTACCTGTGGAGCGCGAGCGGCACCGCAACGAGAACTTCGCGAC	92303
Qy	313	AlaHisAlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaProAlaValPhe	332
Db	92302	ACGCACGCGTGTGCGCAAGATCGCGCGCGCATGGACGCGGAATTCAGAGAACCGCGCTG	92243
Qy	333	PheLysSerGluAlaIleValHisProAspGlnValValGlnTyrIleGlyGln---Asp	351
Db	92242	CTCCTGGCGCAAGCCCAACAGTGGCCCGCAAGACACACAGAAATACCTTCGGCAACGCGGAC	92183
Qy	352	GluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTrpAsnThrLeuAlaThr	371
Db	92182	GATGGCAATGCGCTTCCACTTTCGCTGATCGCGCGCATGTACATGGCGATCGCGCG	92123
Qy	372	ArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHisThr	391
Db	92122	GAGAACCGCTTCCGATCATCCGACATCATGCGCGCAGACCGCGAGGTGCGGCCACCTGC	92063
Qy	392	AlaTrpValAsnTyrValArgSerHisAspAspIleGlyTrpThrPhe---AlaAspGlu	410
Db	92062	CAGTGGCCCATCTCTCTGGCAACACGACGAGCTGACGCTGGAGATGGTGACCGACCGG	92003
Qy	411	AspAlaAlaTyrLeu---GlyIleSerGlyTyrAspHisArgGlnPheLeuAsnArgPhe	429
Db	92002	GAGCGGCACTACCTGTGGGAGGTCTACGCGCAGCGACCGCGCGCGCGGTCAAT-----	91949
Qy	430	PheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsnProSer	449
Db	91949	-----	91949
Qy	450	ThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAlaGlnAspAsp	469
Db	91948	-----CTCGGCATCCGCGCGCGCTGGCGCGCTGTGTGGAGCGCGAC	91907
Qy	470	ProHisAlaValAspArgIleLysLeuLeuTyrSerIleAlaLeuSerThrGlyGlyLeu	489
Db	91906	CGCGCG-----CCGCTGGAGCTGATGAACAGCGCTCTCTTCGTGATGCCCGCGACG	91856
Qy	490	ProLeuIleTyrLeuGlyAspGluValGlyThr-----LeuAsnAspAsp	504
Db	91855	CCGGTATGATTACGGCACGACAGATCGGCATGGCGCAACATCCACTGGCGCGCGCG	91796

Qy	505	Asp	-----TrrpSerGlnAspSerAsn-----	Lysser	513
Db	91795	GAGGGCGTGGCAGCGCCGATGACAGTGTGCCCGACCGCACGGGGCTTCTGGCGCCG		CGCGCC	91736
Qy	514	AspAsp-SerArgTrpAlaHisArgProArgTyr-----	AsnGluAla	527	
Db	91735	GATCGGAGGACAGTGTGTCTGGCGGCATCATGGCTGCTGTACGGCTATGAATCGTTC			91676
Qy	527	aLeuTyr--AlaGlnArgAsnAspProSerThrAlaAlaGlyGlnIleTyrGlnGlyLeu	546		
Db	91675	AACGTGGCGCGCAGACGGCGAC-----	GCGCATTCGTGCTGAATCTGGACG	91628	
Qy	547	ArgHisMetIleAlaValArgGlnSerAsnProArgPheAspGlyGlyArgIleuValThr	566		
Db	91627	CGCGCGTGTGTGCCACGGCGAAGCGCCACCGGTGTGGCGCGCGCAGATCCAGTTC	91568		
Qy	567	PheAsnThrAsnAsnIlyHisIleIleGlyTyrIleArg-----	AsnAsnAla	582	
Db	91567	CTGCAGCGCGCCAAACGAGGTGTGGCCTATGCGCGCGCTGGAAGCGAGGCCGCC	91508		
Qy	583	LeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrVal-----	ThrAlaHis	599	
Db	91507	ATCCTGTGTGGTGGCGAATCTGTCCGGCGCTCGCAGCGCGTGGAGCTGACCTGGCCGCG	91448		
Qy	600	ThrLeuGlnAlaMetProPheIlyAlaHisAspLeuIleGlyIlyThrVal-----	617		
Db	91447	TATGGCAGCGCGTGGCGTG-----	GAGTTGATCGCGCGSACGGCCTTCCCGCG	91397	
Qy	618	-----SerIleuAsnGlnAspLeuThrIleuGlnProTyrGlnIleValMetTrpIleuGluIle	635		
Db	91396	ATCGGCGAGTGGCCCTATCTGTGACGCTCGCGCCCTATGCGCTTCTACTGGCTGAGTGC	91337		

Search completed: November 8, 2003, 21:58:11  
Job time : 5786 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 8, 2003, 17:31:29 ; Search time 413 Seconds  
(without alignments)  
4157.004 Million cell updates/sec

Title: US-09-843-007A-2

Perfect score: 3408

Sequence: 1 MLPTQGVGLIQLVLRIL.....VSLNQDLTLPYQVWMLA 636

Scoring table: BIOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgm2\_1/USPTO.spool\_p/US09843007/runat\_07112003\_170419\_25336/app\_query.fasta\_1.775  
-DB=N\_Geneseq\_19Jun03 -QWTF=fastap -SUFFX=p2n.rng -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNIT5-bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFW=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09843007 -CGEN 1 1 490 @runat\_07112003\_170419\_25336 -NCPU=6 -ICPU=3  
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DSLOP=6 -DELEXT=7

Database : N\_Geneseq\_19Jun03.\*

1: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSI/cgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSI/cgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSI/cgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSI/cgdata/geneseq/geneseq-emb1/NA2002.DAT.\*  
25: /SIDSI/cgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3408	100.0	1939	22	AAF61709	Amylosucrase PCR d
2	3408	100.0	2914	21	AAAI1732	N. polysacchara a
3	3408	100.0	4173	21	AAAO7380	Neisseria polysacc
4	3408	100.0	6878	22	AAF61711	Expression vector
5	3360	98.6	1910	22	AAF61710	Amylosucrase PCR d
6	3360	98.6	6851	22	AAF61712	Expression vector
7	3141.5	92.2	2883	17	AAOT09860	Neisseria polysacc
8	3141.5	92.2	2883	17	AAIT1179	Neisseria polysacc
C 9	1164	34.2	29559	23	AA559346	Propionibacterium
10	560.5	16.4	22934	23	AA559613	Propionibacterium
11	541.5	15.9	3303	23	AA554127	Pseudomonas aerugi
12	529.5	15.5	1794	22	AAH67492	C glutamicum codin
13	529.5	15.5	1981	24	AB565343	DNA encoding C. gl
14	529.5	15.5	34980	22	AAH68531	C glutamicum codin
15	529.5	15.5	34980	22	AAH68532	C glutamicum codin
16	509.5	15.0	4403765	22	AAI99683	Mycobacterium tube
17	509.5	15.0	4411529	22	AAI99682	Mycobacterium tube
18	504	14.8	1704	17	AAIT11413	Sequence encoding
19	488	14.3	3600	17	AAIT13227	Thermotable enzym
20	400.5	11.8	1126	23	AA559731	Propionibacterium
C 21	398.5	11.7	1113	23	AA559702	Propionibacterium
22	393	11.5	1704	24	AEQ67835	Listeria innocua D
23	393	11.5	684707	24	AEQ67196	Listeria innocua D
C 24	393	11.5	3011208	24	AEQ69245	Listeria innocua D
25	375.5	11.0	2944528	24	ABA03041	Listeria innocua D
26	373	10.9	1683	24	AEQ68051	Listeria innocua D
27	373	10.9	1683	24	AEQ70041	Listeria innocua D
28	364	10.7	1102	22	AA711764	Listeria innocua D
C 29	362.5	10.6	29912	23	AA559507	Propionibacterium
C 30	349.5	10.3	2944528	24	ABA03041	Propionibacterium
31	339	9.9	1755	23	ABL26275	Drosophila melanog
32	333.5	9.8	1636	24	ABN66288	Drosophila melanog
33	332.5	9.8	1790	23	ABL26277	DNA encoding alpha
34	331	9.7	1620	25	ABX08486	Staphylococcus aur
35	328.5	9.6	1650	23	AA554857	Drosophila melanog
36	328.5	9.6	1704	23	ABU14333	Staphylococcus aur
C 37	326	9.6	16592	18	AAV74364	Staphylococcus aur
38	324	9.5	4128	23	ABU26276	Drosophila melanog
39	323.5	9.5	1638	23	AA551873	Staphylococcus aur
40	319.5	9.4	1623	24	ABN66287	Streptococcus poly
41	319.5	9.4	2155561	24	ABN71527	Streptococcus poly
C 42	318.5	9.3	1163020	24	ABQ67197	Listeria innocua c
43	317	9.3	1653	25	ABT14986	Pathogen specific
44	317	9.3	1656	22	AAH53337	S. epidermidis ope
45	317	9.3	1683	24	ABN90853	Staphylococcus epi

# ALIGNMENTS

RESULT 1  
AAF61709  
ID AAF61709 standard; DNA; 1939 BP.  
XX  
AC AAF61709;  
XX  
12-JUL-2001 (first entry)  
DT  
DE Amylosucrase PCR derived DNA fragment SEQ ID 4.  
XX

XX Amylosucrase; EC 2.4.1.4; fusion protein; GST; glutathione-S-transferase;  
KW poly(1,4-alpha-glucan); film production; food additive; cyclodextrin; ds.  
XX  
OS Unidentified.  
XX  
FN WO200125449-A2.  
XX

PD 12-APR-2001.  
 XX 04-OCT-2000; 2000WO-EP09695.  
 XX 07-OCT-1999; 99DE-1048408.  
 XX (AXIV-) AXIVA GMBH.  
 XX Bengs H, Polakowski T, Held A, Gallert K;  
 XX WPI; 2001-328330/34.  
 XX Amylosucrase immobilized as fusion protein with anchoring sequence,  
 XX useful in continuous preparation of poly(1,4-alpha-glucan) from sucrose  
 XX  
 XX Claim 3; Page 27-28; 38pp; German.  
 XX This invention describes a novel amylosucrase (AS), immobilizable on a  
 XX solid phase, which comprises a fusion protein (FP) of functional units  
 XX of AS, an anchoring sequence, and optionally additional auxiliary  
 XX sequences. The invention also describes (1) nucleic acid (I) encoding  
 XX FP; (2) expression vector containing (I) and able to express FP in a  
 XX host cell; Escherichia coli containing the vector of (2); (3) anchoring  
 XX sequence, or its functional variants or fragments, of at least 8  
 XX nucleotides that encodes an epitope, a high-affinity binding partner or  
 XX GST (glutathione-S-transferase); (4) solid phase for immobilizing AS  
 XX comprising glutathione-Sepharose; (5) combined, stable catalyst (A)  
 XX comprising FP immobilized on Sepharose for production of (II).  
 XX AS is used in production of poly(1,4-alpha-glucans) (II), useful for  
 XX producing films, as food additives, as starting materials for  
 XX cyclodextrins and as auxiliaries in pharmaceutical formulations.  
 XX Immobilized AS makes possible efficient, inexpensive and continuous  
 XX production of poly(1,4-alpha-glucans) (II), and it can be used  
 XX repeatedly. Compared with known methods, specificity is improved  
 XX (increased yield of (II) and reduced formation of palatinose) and  
 XX reaction is complete within 24 hours, compared to 48-72 hours for batch  
 XX methods. This sequence encodes a PCR derived amylosucrase (EC 2.4.1.4)  
 XX fragment described in the invention.  
 XX  
 XX SQ Sequence 1939 BP; 453 A; 605 C; 482 G; 395 T; 4 other;  
 Alignment Scores:  
 Pred. No.: 0 Length: 1939  
 Score: 3408.00 Matches: 636  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 22 Gaps: 0  
 US-09-843-007A-2 (1-636) x AAF61709 (1-1939)  
 QY 1 MetLeuThrProThrGlnGlnValGlyLeuIleLeuGlnTyrLeuLysThrArgIleLeu 20  
 DB 12 ATGTTGACCCCGCCGAGCAGTCGGTTGATTTTACAGTACCTCAAAACACGCGATCTTG 71  
 QY 21 AspIleTyrThrProGlnGlnArgAlaGlyIleGluLysSerGluAspTrpArgGlnPhe 40  
 DB 72 GACATCTACACGCCCGCCGCGCGCGCGATCGAATAATCCGAGACTGGCGGCGAGT 131  
 QY 41 SerArgArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGly 60  
 DB 132 TCGCGCGCGATGATACGATATTCGCCAACTGATGAGCACTCGACAGCGTGATCGGC 191  
 QY 61 AsnAsnGluAlaLeuLeuProMetLeuGluMetLeuLeuAlaGlnAlaTrpGlnSerTyr 80  
 DB 192 AACAAAGAGCCCTGCTGCTATGCGAATGCTGCTGCGCGAGCATGGCAAGCTAT 251  
 QY 81 SerGlnArgAsnSerSerIleLysAspIleAspIleAlaArgGluAsnAsnProAspTrp 100  
 DB 252 TCCCAAGCGCACTCATCTTTAAAGATATCGATATCGCGCGCGGAAACCAACCCGATGG 311

QY 101 IleLeuSerAsnLysGlnValGlyGlyValCysTyrValAspLeuPheAlaGlyAspLeu 120  
 DB 312 AFTTTGTCCAAACAACAAGTCGGCGGGTGTGCTACGTTGATTGTTGCGCGGCAITTG 371  
 QY 121 LysGlyLeuLysAspLysIleProTyrPheGlnGlnLeuGlyLeuThrTyrIleuHisLeu 140  
 DB 372 AAGGCTTGAAGATAAAATTCCTTATTTTCAAGAGCTTGGTTTGACTTATCTGCACCTG 431  
 QY 141 MetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyr 160  
 DB 432 ATGCGGTGTTAAATGCTCCCTGAAGGCAGGCGCGCTATGCGGTCAGCAGCTAC 491  
 QY 161 ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAlaLeu 180  
 DB 492 CGCGATGTCATTCGCGCACTGGGCACAAATAGCGCACTTGGCGAAGTCATTGCTGCGCTG 551  
 QY 181 HisGluAlaGlyIleSerAlaValAlaValAspPheIlePheAsnHisThrSerAsnGluHis 200  
 DB 552 CACGAGCGCGCATTTCCGCGTGTGCTGATTTATCTTCAACCAACACCTCCACGACAC 611  
 QY 201 GluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePhe 220  
 DB 612 GAATGGCGCAACGCTGCGCGCGCGCGCGCTTTTCGACAAATTTCTACTATATTTC 671  
 QY 221 ProAspArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGln 240  
 DB 672 CCCGACCGCGGATGCGCGCAATACGACGCGACCTGCGGAAATCTTCCCGACCAAG 731  
 QY 241 HisProGlyGlyPheSerGlnLeuGluAspGlyArgTrpValTrpThrPheAsnSer 260  
 DB 732 CACCGCGCGGCTTCTCGCACTGGAAGACGAGCGTGGTGTGCGAGCAGCTTCAATTC 791  
 QY 261 PheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheArgAlaMetAlaGlyGluMet 280  
 DB 792 TTCCAATGGGACTTGAATTCAGCAACCGCGGGGTATTCGCGCAATGGCGGCGAATG 851  
 QY 281 LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTrp 300  
 DB 852 CTGTTCTTCTGCAACTTGGGCGTTGACATCTCTGCTATGATGCGTGGTTCCTTTATGG 911  
 QY 301 LysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPhe 320  
 DB 912 AACCAATGGGACCAAGCTCGCAAAACCTGCGCGAGCGCGACGCCCTCATCGCGCGTTC 971  
 QY 321 AsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHis 340  
 DB 972 AATGCGGTATGCGTATGCGCGCGCGCGCGCTGTTCTTCAATCCGAGGCCATCGTCCAC 1031  
 QY 341 ProAspGlnValValGlnTyrIleGlyGlnAspGlnCysGlnIleGlyTyrAsnProLeu 360  
 DB 1032 CCGGACCAAGTCGTCCAATACATCGGCGAGCAAGATGCCAAATCGGTTACAAACCCCTG 1091  
 QY 361 GlnMetAlaLeuLeuTrpAsnThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAla 380  
 DB 1092 CAATGGGCTTTGTTGGAAACACCTTGCACGCGCGAGTCAACCTGCTCATCAGCGG 1151  
 QY 381 LeuThrTyrArgHisAsnLeuProGluHisThrAlaTrpValAsnTyrValArgSerHis 400  
 DB 1152 CTGACCTACCGCCCAACCTGCGCGAGCATACCGCTGGTCAACTACGTCGCGAGCCAC 1211  
 QY 401 AspAspIleGlyTrpThrPheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyr 420  
 DB 1212 GAGGACATCGGCTGACGCTTTGCGGATGAGACCGCGCATATCTGGGATAGAGCGGCTAC 1271  
 QY 421 AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg 440  
 DB 1272 GACCACCGCAATCTCTCAACCGCTTCTTCGTCACACCGTTTCGACGCGAGCTTCGCTCGT 1331  
 QY 441 GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla 460  
 DB 1332 GCGGTACGTTCCATACACCCAGCACAGCGGACTCCCGGTGTCAGTGGTACAGCCCGG 1391  
 QY 461 AlaLeuValGlyLeuAlaGlnAspAspProHisAlaValAspArgIleLysLeuLeuTyr 480

Db	1392	GCATGGTCGGCTTGGCGCAGACGATCCCCACGCGTGTACCGCATCAAACTTTGTAC	1451
Qy	481	SerIleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr	500
Db	1452	AGCATTCGTTTGAGTACCGCGGCTTCGCGCTGATTTACCTTAGCGCAGAGTGGGTACG	1511
Qy	501	LeuAsnAspAspTyrSerGlnAspSerAsnIlySerAspAspSerArgTyrAlaHis	520
Db	1512	CTCAATGACGACGACTGGTCCGACAGACAGCAATAGCGGACGACGCCGTTGGCGGCAC	1571
Qy	521	ArgProArgTyrAsnGluAlaIleuTyrAlaGlnArgAsnAspProSerThrAlaAlaGly	540
Db	1572	CGTCGGCGCTACACGACGAGCCCTGTACGCGACGACGACGATCCGTCGACGCGACCGGG	1631
Qy	541	GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp	560
Db	1632	CAAACTCATCAGGCGTTGGCGCATATGATTCGCGTCGCCCAAGCAATCCGCGCTTCGAC	1691
Qy	561	GlyGlyArgLeuValThrPheAsnThrAsnAsnIlyHisIleIleGlyTyrIleArgAsn	580
Db	1692	GGCGCGAGCGCTGGTTTACATTCACACCAACACGACGACATCATCTCGCTCATCTCCGCAAC	1751
Qy	581	AsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThr	600
Db	1752	AATCGCGCTTTGGCATTCGGTAATCTCAGCGAAATATCCGCAACCGTTACCGCGCATACC	1811
Qy	601	LeuGlnAlaMetProPheIlyeAlaHisAspLeuIleGlyGlyysThrValSerLeuAsn	620
Db	1812	CTGCAAGCCATGCCCTTCAGCGCGACGACCTCATCGGTGCMAAATCTGTACGCTGAAT	1871
Qy	621	GlnAspLeuThrLeuGlnProTyrGlnValMetTyrLeuGluIleAla	636
Db	1872	CAGGATTTGAGCTTCAGCCCTATCAGCTCATGTGGCTCGAAATTCGCC	1919

RESULT 2

RESULT 2  
AAA11732

AA11732  
ID AA11732 standard; DNA: 2914 bp.

XX  
ID 1732

AA  
AC  
BBB11732.

AC  
VV  
AAAI/32;

3  
 4  
 5  
 6

DT 21-JUL-20



DE N. polysa

XX

KW Alpha-1,6

101 KW  
transpenni  
transpenni

**KW** ultraviolet  
**transgen**

ADJUTANT GENERAL  
U.S. ARMY  
WASHINGTON, D.C.

any loss of  
KY

33

OS Neisseria



**FF: Key**

FT CDS

33

$$\frac{1}{2} \frac{d}{dt} \left( \frac{1}{2} \frac{d}{dt} \right)$$
$$\mathbf{F}_1 = \mathbf{Y}$$

XX

XX New nucleic acid encoding a branching enzyme, useful for in vitro  
PT synthesis of branched glucans and to prepare transgenic plants  
PT producing modified starch -  
XX  
PS Disclosure; Page 99-102; 115pp; German.  
XX  
CC This invention describes a novel nucleic acid (I) isolated from  
CC *Neisseria* which encodes a branching enzyme (II). (I) is used for  
CC recombinant production of (II) subsequently used in the in vitro  
CC production of alpha-1,6-branched alpha-1,4-glucans. It is also used to  
CC prepare transgenic plants that produce starches with modified properties.  
CC (II) are used as binders for tablets, carriers for pharmaceuticals,  
CC flavors and perfumes and powdered additives, packaging materials,  
CC ultra-violet light adsorbers in sunscreens and also for any of the usual  
CC applications of starch in foods, papermaking, as textile size, in soil  
CC stabilization, as wetting agent for agricultural chemicals, as polymer  
CC additives etc. Fragments of (I) are useful as PCR primers and antisense  
CC molecules or ribozymes for inhibiting expression of (I), and the  
CC regulatory region of (II) can be used to control expression of  
CC heterologous sequences in host cells. (I) provides an inexpensive method  
CC for producing alpha-1,6-branched alpha-1,4-glucans (III), producing  
CC products that can be tailored for particular applications, particularly  
CC by controlling the degree of branching. Starch from transgenic plants  
CC has increased gel strength; reduced phosphate content; reduced peak  
CC viscosity, lower pasting temperature and granule size and/or altered  
CC starch chain distribution. This sequence encodes an amylosucrase isolated  
CC from *Neisseria polysacchara* which is described in the method of the  
CC invention.

SQ Sequence 2914 BP; 662 A; 868 C; 773 G; 611 T; 0 other;

**Alignment Scores:**

Pred. No.:

FILE NO.: 3408-00  
SCORE:

Percent Similarity: 100.00%

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 100.00%

DB: 21

alignment scores:		
Pred. No.:	0	2914
Length:		

Field: no.:	0	Length:	2914
Score:	3408.00	Matches:	636

score:	3408.00	636
Percent similarity:	100.00%	0
Conservative:		
MacCrees:		

Percent similarity:	100.00%	Conservative:	0
Best local similarity:	100.00%	Conservative:	0

```
Best Local Similarity: 100.00%
Mismatches: 0
```

Query Match: 100.00% Indels: 0

US-09-843-007A-2 (1-636) x AAA11732 (1-2914)

Qy	1	MetLeuThrProThrGlnGlnValGlyLeuIleLeuGlnTyrLeuLysThrArgIleLeu	20
Db	957	ATGTGTGACCCCAACGACGACGTCGGTTTGATTTTACAGTACCTCAAAACACGCACTCTTG	1016
Qy	21	AspIleTyrThrProGluGlnArgCAlaGlyIleGluLysSerGluAspTrpAraGlnPhe	40
Db	1017	GACATCTACACGCCCAAGACAGCGCCGGCATCGAAATAATCCGAAGACTGCGCGCAGTTT	1076
Qy	41	SerArgArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGly	60
Db	1077	TCGGCGCGCATGATACGCATTTCCCCAAACTCATGAACCAACTCGACACGCGTGTACGGC	1136
Qy	61	AsnAsnGluAlaLeuLeuProMetLeuGluMetLeuLeuAlaGlnAlaTrpGlnSerTyr	80
Db	1137	AACAACGAAGCCCTGCTGCCCTATGCTGGAAATGCTGCTGGCGCAGGCATGGCAAAAGCTAT	1196
Qy	81	SerGlnArgAsnSerSerLeuLysAspIleAspIleAlaArgGluAsnAsnProAspTyr	100
Db	1197	TCCCAACGCAACTCATCTTTAAAGATATCGATATCGCGCGCAAAAACAACCCCGATTGG	1256
Qy	101	IleLeuSerAsnLysGlnValGlyGlyValCysTyrValAspLeuPheAlaGlyAspLeu	120
Db	1257	ATTTTGTGCCAACAAACAAAGTCGCGCGCGTGTGCTAGCTTGATTTGTTGTGCGCGCATTTG	1316
Qy	121	LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeu	140
Db	1317	AAGGGCTTGAAGATAAATTCCTTATTTTCAGAGCTTGGTTTGACTTATCTATCTGCACCTG	1376
Qy	141	MetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyr	160
Db	1377	ATCCCGCTGTTTAAATGCCCTCAAGCGAAAGCGACGGCGGCTATCGGTTCACGAGCTAC	1436

QY 161 ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaLeu 180  
 Db 1437 CGCGATGTCATTCGCGACATGGGACATAGGACATTCGCGAGATCATTCGCGCTG 1496  
 QY 181 HisGluAlaGlyIleSerAlaValAlaAspPheIlePheAsnHisThrSerAsnGluHis 200  
 Db 1497 CAGAGCGCGCATTCGCGCGCTGCTGATTTATCTCAACACACACCTCCCAACGACAC 1556  
 QY 201 GluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTrpIlePhe 220  
 Db 1557 GAATGGGCGCAACGCTGCGCGCGCGCGACCGCTTTCGACAAATTCCTACTATATTTTC 1616  
 QY 221 ProAspArgMetProAspGlnTyrAspArgThrIleuArgGluIlePheProAspGln 240  
 Db 1617 CCGAGCGCGCGATGCGCGACCAATACGACGCGACCGCTGCGGAAATCTTCGCGGACGAG 1676  
 QY 241 HisProGlyGlyPheSerGlnLeuGluAspGlyArgTrpValTrpThrPheAsnSer 260  
 Db 1677 CACCGCGCGGCTTCTCGCACTTGAAGACGACGCTGGGTGTGGACGACCTTCAATTC 1736  
 QY 261 PheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheArgAlaMetAlaGlyGluMet 280  
 Db 1737 TTCGAATGGGACCTGAATACAGCAACCCGCTGGGTATTCGCGCAATGGCGGCGAAATG 1796  
 QY 281 LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTrp 300  
 Db 1797 CTGTTCCTCCCACTTGGCGGTGACATCTCGGTATGATCGGTGCTTTATTGG 1856  
 QY 301 LysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPhe 320  
 Db 1857 AAACAAATGGGACAGCTCGGAAACCTCGCGACGCGCACCGCTCATCGCGGCTTC 1916  
 QY 321 AsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHis 340  
 Db 1917 AATGCCGTATGCTATGCTATGCGCGCGCGCTGCTTCTCAATCCGAGCGACATCGTCCAC 1976  
 QY 341 ProAspGlnValValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnProLeu 360  
 Db 1977 CCGGACCACTCTGCCAATACATCGGCGAGGACGAATGCGCAATCGTTACACCCCTCG 2036  
 QY 361 GlnMetAlaLeuLeuTrpAsnThrLeuAlaThrArgGluValAsnLeuHisGlnAla 380  
 Db 2037 CAAATGGCATTTGTGGAAACACCTTCGCCACGCGGAGTCAACCTGCTCCATCAGGCG 2096  
 QY 381 LeuThrTyrArgHisAsnLeuProGluHisThrAlaTrpValAsnTyrValArgSerHis 400  
 Db 2097 CTGACCTACCGCCACACACCTTCCCGAGCATACCGCTGGGTCAACTACGTCGCGACCCAC 2156  
 QY 401 AspAspIleGlyTrpThrPheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyr 420  
 Db 2157 GAGGACATCGCTGGACGTTTGGCGATGAAGACGCGCATATCTGGGCATAAGCGCTAC 2216  
 QY 421 AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg 440  
 Db 2217 GACACCGCGCAATTCCTCAACCGCTTCTCGTCAACCGTTTCGACGCGAGCTTCGCTCGT 2276  
 QY 441 GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla 460  
 Db 2277 GCGGTACCGTTCCATACACCCACGACGACGCGCATGCGGTGTCACTGGTACGCGCGG 2336  
 QY 461 AlaLeuValGlyLeuAlaGlnAspAspProHisAlaValAspArgIleLysLeuLeuTyr 480  
 Db 2337 GCATTGGTGGCTTGGCGAAGACGATCCCGCGCGCTTGACCGCATCAAACTCTTGTAC 2396  
 QY 481 SerIleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr 500  
 Db 2397 AGCATTTGCTTTAGTACCGCGCTTCTCGCTGATTTTACTAGCGACGAGTGGGTACG 2456  
 QY 501 LeuAsnAspAspTrpSerGlnAspSerAsnLysSerAspAspSerArgTrpAlaHis 520  
 Db 2457 CTCAATGACGACGCTGCTCGAAGACAGCAATAGAGCGACGACGACGCTGGCGGCGAC 2516

QY 521 ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGly 540  
 Db 2517 CGTTCGCGCTACACGAGCCCTGTATCGGCGAACGACGACGATCCGTGACCGACGCGG 2576  
 QY 541 GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp 560  
 Db 2577 CAAATCTATCATGGGCTTGGCCATATGATTCGCTCGCGCAAGCAATCGCGCTTCGAC 2636  
 QY 561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArgAsn 580  
 Db 2637 GCGGCGAGGCTGGTATCATTCACCAACCAACGACATCATCGCTACATTCGCAAC 2696  
 QY 581 AsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThr 600  
 Db 2697 AATGCGCTTTTGGCATTCGTAACCTTCAGCGAATATCGCAACCGTTACCGCGCATAC 2756  
 QY 601 LeuGlnAlaMetProPheIleAlaHisAspIleIleGlyGlyValThrValSerLeuAsn 620  
 Db 2757 CTGCAAGCATGCCCTTCAGGCGACGACCTCATCGGTGGCAAACTGTTCAGCTGAAT 2816  
 QY 621 GlnAspLeuThrLeuGlnProTyrGlnValMetTrpIleuGluIleAla 636  
 Db 2817 CAGGATTTGACGCTTCAGCCCTATCAGGTCTAGTGTGCTCGAAATCGCC 2864  
 RESULT 3  
 AAA07380  
 ID AAA07380 standard; DNA; 4173 BP.  
 AC AAA07380;  
 XX 30-JUN-2000 (first entry)  
 DT  
 XX  
 DE Neisseria polysaccharea amylosucrase coding sequence.  
 XX  
 KW Amylosucrase; linear alpha-1,4 glucan production; fructose production;  
 extracellular enzyme; cyclodextrin production; pure fructose syrup; ds.  
 OS Neisseria polysaccharea.  
 XX  
 PN WO200014249-A1.  
 PD 16-MAR-2000.  
 XX  
 PF 02-SEP-1998; 98WO-EP05573.  
 XX  
 PR 02-SEP-1998; 98WO-EP05573.  
 XX  
 PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.  
 XX  
 PI Quanz M, Provart N;  
 XX  
 DR WPI: 2000-256995/22.  
 DR P-PSDB; AAY81945.  
 XX  
 PT Novel amylosucrase polynucleotides and polypeptides obtained from  
 Neisseria polysaccharea, used for production of alpha-1,4 glucans,  
 fructose and cyclodextrins -  
 XX  
 PS Claim 1; Page 25-30; 36pp; English.  
 XX  
 CC This sequence encodes the Neisseria polysaccharea amylosucrase of the  
 invention. The amylosucrase sequences are used for the in vitro  
 production of linear alpha-1,4 glucans and fructose, as the enzyme is an  
 extracellular enzyme. Amylosucrases may also be useful for the production  
 of cyclodextrins. The amylosucrase sequences of the invention allow for  
 the in vitro production of alpha-1,4 glucans and pure fructose syrup at  
 low cost. Activated glucose derivatives or cofactors are not required.  
 CC Immobilised cells can be at much higher densities than cells in liquid  
 cultures, resulting in higher productivity.  
 XX  
 SQ Sequence 4173 BP; 926 A; 1196 C; 1118 G; 933 T; 0 other;  
 Alignment Scores:

Pred. No.: 0 Length: 4173  
 Score: 3408.00 Matches: 636  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 21 Gaps: 0

US-09-843-007a-2 (1-636) x AAA07380 (1-4173)

QY	1	MetLeuThrProThrGlnGlnValGlyLeuLeuLeuGlnTyrLeuLeuThrArgLeuLeu	20
DB	1971	ATGTTGACCCCGCAGCAGTGGTGGTTGATTTACAGTACTCTCAAAACAGCATCTTG	2030
QY	21	AspLeuThrProGlnArgGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln	40
DB	2031	GACATCTACAGCCCGCAACAGCGCGCGCATCGAAATCCGAGAGCTGGCGGCGATTT	2090
QY	41	SerArgArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGly	60
DB	2091	TCGGCGCGCATGATACGCATTTCCCAACCTGATGAACGACTCGACAGCGGTGTACGGC	2150
QY	61	AsnAsnGluAlaLeuLeuProMetLeuGluMetLeuLeuAlaGlnAlaTyrGlnSerTyr	80
DB	2151	AACACAGAGCCCTGCTGCTATGCTGAAATGCTGCTGGCGCAGCATGGCAAGATAT	2210
QY	81	SerGlnArgAsnSerSerLeuLysAspIleAspIleAlaArgGluAsnAsnProAspTyr	100
DB	2211	TCCCAACGCACTCATCTTAAAGATATCGATATCGCGCGGAAACCAACCCCGATTGG	2270
QY	101	IleLeuSerAsnLysGlnValGlyValCysTyrValAspLeuPheAlaGlyAspLeu	120
DB	2271	ATTTTGTCCAAACAAAGTGGCGGCGTGTGCTAGCTTGTATTTGTTGCGCGCATTTG	2330
QY	121	LysGlyLeuLysAspLysIleProTyrPheGlnGlnGlnGlnGlnGlnGlnGlnGln	140
DB	2331	AAGGCTTGAAGATAAAATTCCTATTTTCAAGAGCTTGTGTTGACTTATCTGCACCTG	2390
QY	141	MetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyr	160
DB	2391	ATGCGCGTGTAAATGCTTGAAGCAAAAGCGAGCGCGCTATGCGGTTCAGCAGCTAC	2450
QY	161	ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaLeu	180
DB	2451	CGCGATGTCAATCCGCGCATCGGCGACCAATAGCGACTTGGCGGAAGTCATTGCTGCGCTG	2510
QY	181	HisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSerAsnGluHis	200
DB	2511	CACGAGCGCGCATTTCCGCGCTCGTGGATTTTATCTTCAACACCACTCCCAACGAACAC	2570
QY	201	GluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePhe	220
DB	2571	GAATGGCGCAACGCTGCGCGCGCGCGACCCGCTTTTCGACAAATTTCTACTATATTTTC	2630
QY	221	ProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGln	240
DB	2631	CCCGACCGCGGATGCGCGACCAATAGCGACCGACCTGGCGGAATCTTCCCGCGACCG	2690
QY	241	HisProGlyGlyPheSerGlnLeuGluAspGlyArgTyrValTyrThrThrPheAsnSer	260
DB	2691	CACCGCGCGGCTTCTCGCACTGGAAGACGCGCGCTGGGTGGACGACCTTCAATTC	2750
QY	261	PheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheArgAlaMetAlaGlyGluMet	280
DB	2751	TTCCAAATGGGACCTTGAATACAGCAACCGCGGATATTCGCGCAATGGCGGCGGAAATG	2810
QY	281	LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTyr	300
DB	2811	CTGTTCTTGCACACTTGGGCGTTGACATCTGGGTATGATGCGGTGCTTTATTTGG	2870
QY	301	LysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPhe	320
DB	2871	AAACAAATGGGGACCAAGCTGCAAAACCTGCGCGAGCGCGACGCGCTCATCCGCGCGTTC	2930

## RESULT 4

AAP61711  
 ID AAP61711 standard; DNA; 6878 BP.  
 XX AAP61711;  
 AC AAP61711;  
 XX  
 DT 12-JUL-2001 (first entry)  
 XX  
 DE Expression vector pGEX-4T-1-AmsU containing amylosucrase DNA.

QY	321	AsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHis	340
DB	2931	AATGCGGTATGCGGTATGCGCGCGCGCGGTCTTCAATCGGAAGCCATCGTCCAC	2990
QY	341	ProAspGlnValValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnProLeu	360
DB	2991	CCCGACCAAGTCGTCCTCAATACATCGGCGAGCAGCAATGCCAAATCGGTTCACACCCCTG	3050
QY	361	GlnMetAlaLeuLeuTyrAsnThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAla	380
DB	3051	CNAATGGCATGTTGTGGACACCCCTTGCACCGCGCGAAGTCAACCTGCTCCATCAGCGC	3110
QY	381	LeuThrTyrArgHisAsnLeuProGluHisThrAlaTyrValAsnTyrValArgSerHis	400
DB	3111	CTGACCTACCGCCACCACTGCGCGAGCATACCGCTGGGTCAACTACGTACGCGAGCCAC	3170
QY	401	AspAspIleGlyTyrThrPheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyr	420
DB	3171	GACGACATCGCGTGGAGCTTTCGATGAAGACGCGGCATATCTGGGCGATAAGCGGCTAC	3230
QY	421	AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg	440
DB	3231	GACCAACCGCAATTCCTCAACCGCTTCTCGTCAACCGTTTCGACGCGCAGCTTCGCTCGT	3290
QY	441	GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla	460
DB	3291	GCGTACCGTTCCAATCAACCAACCAAGCAGCGGCGACTGCGGTGTCACTGTGTACAGCGCG	3350
QY	461	AlaLeuValGlyLeuAlaGlnAspAspProHisAlaValAspArgIleLysLeuLeuTyr	480
DB	3351	GCATTGCTGCGCTTGGCGCAAGCAGTCCCGACCGCTTGACCGCATCAACTCTTGCTAC	3410
QY	481	SerIleAlaLeuSerThrGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr	500
DB	3411	AGCATTTGCTTTGAGTACCGCGCTCTGCGCTGATTACCTAGGCGCAGGAAGTGGTACG	3470
QY	501	LeuAsnAspAspAspTyrSerGlnAspSerAsnLysSerAspAspSerArgTyrAlaHis	520
DB	3471	CTCAATGACGACGACCTTGGTGGCAGACAGCAGCATTAAGAGCAGCAGCGCGCTGGCGCGC	3530
QY	521	ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGly	540
DB	3531	CGTCCCGGTACCAACCAAGCCCTGTACGCGCAACGACGATCCGTCGACCGCGCGG	3590
QY	541	GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp	560
DB	3591	CAATCTATCAGGCGCTTGGCGCATATGATTTGCGTGGCGCAAGCAATCGCGCTTCGAC	3650
QY	561	GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArgAsn	580
DB	3651	GGCGGCGAGCTGTGTACATTTCAACCAACCAACGACATCATCGCTACATCCGCAAC	3710
QY	581	AsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThr	600
DB	3711	AATGCGCTTTTGGCATTCGTTAACTTCAGCGAATATTCGCAAAACCGCTTACCGCGCATCC	3770
QY	601	LeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyLysThrValSerLeuAsn	620
DB	3771	CTGCAAGCCATGCCCTTTCAGGCGGACGACCTCATCGGTGGCGCAAAACTGTGACGCTGAAT	3830
QY	621	GlnAspLeuThrLeuGlnProTyrGlnValMetTyrLeuGluIleAla	636
DB	3831	CAGGATTTGACGCTTCAGCGCTTATCAGGTCTATGCTGGCTCGAAATCGCC 3878	





QY 421 AspHisArgGlnPheLeuAsnArgPheValAsnArgPheAspGlySerPheAlaArg 440  
 Db 2205 GACCACCGCAATTCCTCAACCGCTCTTCGTCAACCGTTTCGACGGAGCTTCGCTCGT 2264  
 QY 441 GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla 460  
 Db 2265 GCGTACCGCTTCATATACAAACCAACACAGCGCACTGCGGTGTCAGTGGTACAGCCGG 2324  
 QY 461 AlaLeuValGlyLeuAlaGlnAspAspProHisAlaValAsnArgGlyLeuValGlyThr 480  
 Db 2325 GCATGGTTCGGTTCGGCGAAGCGATCCCGCGGTGACCGCATCAAACTCTTGAC 2384  
 QY 481 SerIleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr 500  
 Db 2385 AGCAATTCGTTCAGTACCGCGCTTCGCGCTGATTTACCTAGGCGACGAAGTGGTACG 2444  
 QY 501 LeuAsnAspAspAspTyrSerGlnAspSerAsnLysSerAspAspSerArgTrpAlaHis 520  
 Db 2445 CTCATATGACGACACTGGTTCGAAGACACCAATAGAGCGACAGACCGGTGGCGGAC 2504  
 QY 521 ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGly 540  
 Db 2505 CGTCGCGCTACAAAGCAAGCCCTGTACGCGCAACGCAACGATCCGTTCGACCGCGG 2564  
 QY 541 GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp 560  
 Db 2565 CAAATCTATACGGGCTTCGCGCATATGATTCGCGTCCGCGCAAGCAATCCGCGCTTCAC 2624  
 QY 561 GlyGlyArgLeuValThrPheAsnThrAsnLysHisIleIleGlyTyrIleArgAsn 580  
 Db 2625 GCGCGCAGGCTGGTTACATTTCAACACCAACACACAGCACATCATCGGTACATCCGCAAC 2684  
 QY 581 AsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThr 600  
 Db 2685 AATGCGCTTTTGGCATTCGGTAACCTTCAGCGAATATCCGCAACCGTTACCGCGCATACC 2744  
 QY 601 LeuGlnAlaValProPheLysAlaHisAspLeuIleGlyGlyLysThrValSerLeuAsn 620  
 Db 2745 CTGCAAGCCATGCCCTTCAGCGCGACGACCTTCATCGGTGGCAAACTGTACGCTGAAT 2804  
 QY 621 GlnAspLeuThrLeuGlnProTyrGlnValMetTrpLeuGluIleAla 636  
 Db 2805 CAGGATTCGACCTTCAGCCCTATCAGGTTCATGTGGCTCGAATCGCC 2852

RESULT 5  
 AAF61710  
 ID AAF61710 standard; DNA; 1910 BP.  
 AC AAF61710;  
 DT 12-JUL-2001 (first entry)  
 XX  
 DE Amylosucrase PCR derived DNA fragment AmSu5.  
 XX  
 KW Amylosucrase; EC 2.4.1.4; fusion protein; GST; glutathione-S-transferase;  
 XX poly(1,4-alpha-glucan); film production; food additive; cyclodextrin; ds.  
 OS Unidentified.  
 XX  
 FN W0300125449-A2.  
 XX  
 PD 12-APR-2001.  
 XX  
 PF 04-OCT-2000; 2000WO-BP09695.  
 XX  
 PR 07-OCT-1999; 99DE-1048408.  
 XX  
 PA (AXIV-) AXIVA GMBH.  
 XX  
 PI Bengs H, Polakowski T, Held A, Gallert K;  
 XX  
 DR WPI; 2001-328330/34.  
 XX

PT Amylosucrase immobilized as fusion protein with anchoring sequence,  
 PT useful in continuous preparation of poly(1,4-alpha-glucan) from sucrose  
 XX  
 PS Claim 3; Page 28; 38pp; German.  
 XX  
 CC This invention describes a novel amylosucrase (AS), immobilizable on a  
 CC solid phase, which comprises a fusion protein (FP) of functional units  
 CC of AS, an anchoring sequence, and optionally additional auxiliary  
 CC sequences. The invention also describes (1) nucleic acid (I) encoding  
 CC host cell; Escherichia coli containing (I) and able to express FP in a  
 CC culture; or its functional variants or fragments, of at least 8  
 CC nucleotides that encodes an epitope, a high-affinity binding partner or  
 CC GST (glutathione-S-transferase); (4) solid phase for immobilizing AS  
 CC comprising glutathione-Sepharose; (5) combined, stable catalyst (A)  
 CC comprising FP immobilized on Sepharose for production of  
 CC poly(1,4-alpha-glucan) (II); and (6) biocatalytic production of (II).  
 CC AS is used in production of poly(1,4-alpha-glucans) (II), useful for  
 CC producing films, as food additives, as starting materials for  
 CC cyclodextrins and as auxiliaries in pharmaceutical formulations.  
 CC Immobilized AS makes possible efficient, inexpensive and continuous  
 CC production of poly(1,4-alpha-glucans) (II), and it can be used  
 CC repeatedly. Compared with known methods, specificity is improved  
 CC (increased yield of (II) and reduced formation of palatinose) and  
 CC reaction is complete within 24 hours, compared to 48-72 hours for batch  
 CC methods. This sequence encodes a PCR derived amylosucrase (EC 2.4.1.4)  
 CC AmSu5 fragment described in the invention.  
 XX  
 SQ Sequence 1910 BP; 447 A; 595 C; 474 G; 387 T; 7 other;

## Alignment Scores:

Pred. No.: 0 Length: 1910  
 Score: 3360.00 Matches: 626  
 Percent Similarity: 100.00% Conservative: 1  
 Best Local Similarity: 99.84% Mismatches: 0  
 Query Match: 98.59% Indels: 0  
 DB: 22 Gaps: 0

US-09-843-007A-2 (1-636) x AAF61710 (1-1910)

QY 10 LeuIleLeuGlnTyrLeuLysThrArgIleLeuAspIleTyrThrProGluGlnAla 29  
 Db 10 ATGATTTTACAGTACTCTCAAAACACGCGATCTTGGACATCTACACCGCGACAGCGGCC 69  
 QY 30 GlyIleGluLysSerGluAspTrpArgGlnPheSerArgArgMetAspThrHisPhePro 49  
 Db 70 GGCATCGAAAATCCGAAGACTCGCGCGCAGTTTTCGCGCGCGCATGGATCGCATTTCCCC 129  
 QY 50 LysLeuMetAsnGluLeuAspSerValTyrGlyAsnAsnGluAlaLeuLeuProMetLeu 69  
 Db 130 AACTGATGATGACGNACTCGACGCGTGTACGGCAACACGACGCGCTGTGCTATGCTG 189  
 QY 70 GluMetLeuLeuAlaGlnAlaTrpGlnSerTyrSerGlnArgAsnSerSerLeuLysAsp 89  
 Db 190 GAATATGCTGTGGCGCAGGCGATGCGAAAGCTATTCCACGCGCAACTCATCTCTTAAAGAT 249  
 QY 90 IleAspIleAlaArgGluAsnAsnProAspTrpIleLeuSerAsnLysGlnValGlyGly 109  
 Db 250 ATCGATATCGCGCGCGAATAACACCCGATGGATTGTTTCCCAACAAACAGTCGCGCGC 309  
 QY 110 ValCysTyrValAspLeuPheAlaGlyAspLeuLysGlyLeuLysAspLysIleProTyr 129  
 Db 310 GTGTGCTACCTTCATTTGTTTTCGCGCGATTTGAAGGCGCTTGAAGAGATAAAATTCCTTAT 369  
 QY 130 PheGlnGluLeuGlyLeuThrTyrLeuHisLeuMetProLeuPheLysCysProGluGly 149  
 Db 370 TTTCAGAGCTTTGGTATTGACTTATCTGACCTGATCGCGCTGTGTTAAATGCCCTGAAGGC 429  
 QY 150 LysSerAspGlyGlyTyrAlaValSerSerTyrArgAspValAsnProAlaLeuGlyThr 169  
 Db 430 AAAAGCGACGCGGCTATGCGGTACGAGCTACCGCGATCTCAATCCGCACTGGGCACA 489

```
QY 170 IleGlyAspLeuArgGluValIleAlaLeuHisGluAlaGlyIleSerAlaValVal 189
DB 490 ATAGGGGACTTGGCGAAGTCATGTGCTGGCTGCACGAAGCCGCGCATTTCCGCGCTGTC 549
QY 190 AspPheIlePheAsnHisThrSerAsnGluHisGluTrpAlaGlnArgCysAlaAlaGly 209
DB 550 GATTTTATCTTCAACCACTTCCACGACCTCAACGACAGCAATGGCGCAAGCTTGGCGCGCGC 609
QY 210 AspProLeuPheAspAsnPheTyrTrpIlePheProAspArgArgMetProAspGlnTyr 229
DB 610 GACCCGCTTTTCGACAATTTCTACTATATTTTCCCGACCGCGGATGCGCCGACCAATAC 669
QY 230 AspArgThrLeuArgGluIlePheProAspGlnHisProGlyGlyPheSerGlnLeuGlu 249
DB 670 GACCGCACCTTGGCGCAATCTTCCCGACACGACCGCGCGCTTCTCGCACTGGAA 729
QY 250 AspGlyArgTrpValTrpThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsn 269
DB 730 GACGGACGCTGGGTGTGGACGACCTTCAATTCCTTCCAAATGGGACTTGAATTACAGCAAC 789
QY 270 ProTrpValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAsp 289
DB 790 CCGTGGGTATTCGCGCAATGGCGGCGAAATGCTGTTCTTGCCAACTTGGCGGTGAC 849
QY 290 IleLeuArgMetAspAlaValAlaPheIleTrpLysGlnMetGlyThrSerCysGluAsn 309
DB 850 ATCTCGGTATGGATCGGTGCTTTATTTGGAAACAATGGGACAGCTGGCGAAAC 909
QY 310 LeuProGlnAlaHisAlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaPro 329
DB 910 CTGCGCAGCGGACGACCCCTCACTCCGCGGTTCATGCGGTATGCGTATGCGCGGCC 969
QY 330 AlaValPhePheLysSerGluAlaIleValHisProAspGlnValValGlnTyrIleGly 349
DB 970 GCGGTCTTCTCAAATCCGAGCAGCATGCTCCACCCGACCAAGTGTCCAATACATCGG 1029
QY 350 GlnAspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTrpAsnThrLeu 369
DB 1030 CAGGACGAATGCCAAATCGGTTACAAACCCCTGCAAAATGGCATTTGTGGAAACCCCTT 1089
QY 370 AlaThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGlu 389
DB 1090 GCCACGCGGAGTCAACTGTCTCATCAGCGCTGACTACCGCACCAACTGCGCGAG 1149
QY 390 HisThrAlaTrpValAsnTyrValArgSerHisAspAspIleGlyTrpThrPheAlaAsp 409
DB 1150 CATACGCGCTGGGTCAACTACGTCGCGCAGCAGCAGCATCGGCTGGACGTTTGGCGAT 1209
QY 410 GluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsnArgPhe 429
DB 1210 GAAGACGCGGCATATCTGGGCATAAGCGGTACGACCAACCGCCAAATTCCTCAACGCTTC 1269
QY 430 PheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsnProSer 449
DB 1270 TTGCTCAACGTTTCGACGCGACGTTTCGCTCGTGGGTACCGTTCCTCAACCAACCAAGC 1329
QY 450 ThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAlaGlnAspAsp 469
DB 1330 ACAGGCGACTCGCGTGCAGTGATACAGCGCGGCATTTGGTGGCTTGGCGCAAGACGAT 1389
QY 470 ProHisAlaValAspArgIleTyrLeuLeuTyrSerIleAlaLeuSerThrGlyGlyLeu 489
DB 1390 CCCCACGCGGTTCACCGCATCAACATCTTGTACAGCATTTGCTTGGATACCGGCGTCTG 1449
QY 490 ProLeuIleTyrLeuGlyAspGluValGlyThrLeuAsnAspAspTrpSerGlnAsp 509
DB 1450 CCGCTGATTTACCTAGCGCACGAAGTGGGTGATGCTCAATCAGCACGACGCTGGTGGCAAGAC 1509
QY 510 SerAsnLysSerAspAspSerArgTrpAlaHisArgProArgTyrAsnGlnAlaLeuTyr 529
DB 1510 AGCAATATAGAGCGACGACGCGGTGGCGCACCGTCCGCGCTACCAACGAGCCCTGTAC 1569
QY 530 AlaGlnArgAsnAspProSerThrAlaAlaGlyGlnIleTyrGlnGlyLeuArgHisMet 549
```

---

```
DB 1570 GCGCAACCAACGATCCGTCGACCGCGCAAAATCTATCAGGCGTTCCGCGCATATG 1629
QY 550 IleAlaValArgGlnSerAsnProArgPheAspGlyGlyArgLeuValThrPheAsnThr 569
DB 1630 ATTGCGGTCCGCGCAAGCAATCCGCGCTTCACGCGCGGCGGCTGGTTACATTCACAC 1689
QY 570 AsnAsnLysHisIleIleGlyTyrIleArgAsnAsnAlaLeuLeuAlaPheGlyAsnPhe 589
DB 1690 AACCAACAGCACATCATCGGTACATCCGCAACCAATCGCTTTTGGCATTCGGTAAC 1749
QY 590 SerGluTyrProGlnThrValThrAlaHisThrLeuGlnAlaMetProPheLysAlaHis 609
DB 1750 AGCGAATATCCGCAACCGTTACCGCGATACCTTCGACGACATGCCCTTCAGGCGCAC 1809
QY 610 AspLeuIleGlyGlyLysThrValSerLeuAsnGlnAspLeuThrLeuGlnProTyrGln 629
DB 1810 GACCTCATCGGTGGCAAACTGTGACGCTGAATCAGGATTTGACGCTTCAGCCCTATCAG 1869
QY 630 ValMetTrpLeuGluIleAla 636
DB 1870 GTCATGTGCTCGAAATCGCC 1890

RESULT 6
AAF61712
ID AAF61712 standard; DNA; 6851 BP.
XX
AC AAF61712;
XX
DT 12-JUL-2001 (first entry)
XX
DE Expression vector pGEX-4T-1-AmsU5 containing amylosucrase DNA.
XX
KW Amylosucrase; EC 2.4.1.4; fusion protein; GST; glutathione-S-transferase;
XX poly(1,4-alpha-glucan); film production; food additive; cyclodextrin; ds.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 258..2828
FT /*tag= a
FT /product= "AmsU5"
XX
FN WO200125449-A2.
XX
PD 12-APR-2001.
XX
PF 04-OCT-2000; 2000WO-EP09695.
XX
PR 07-OCT-1999; 99DS-1048408.
XX
PA (AXIV-) AXIVA GMBH.
XX
PI Bengs H, Polakowski T, Held A, Gallert K;
XX
DR WPI: 2001-328330/34.
DR P-PSDB; AAB570883.
PT Amylosucrase immobilized as fusion protein with anchoring sequence,
PT useful in continuous preparation of poly(1,4-alpha-glucan) from sucrose
PS Claim 6; Page 31-34; 38pp; German.
XX
CC This invention describes a novel amylosucrase (AS), immobilizable on a
CC solid phase, which comprises a fusion protein (FP) of functional units
CC of AS, an anchoring sequence, and optionally additional auxiliary
CC sequences. The invention also describes (1) nucleic acid (I) encoding
CC FP; (2) expression vector containing (I) and able to express FP in a
CC host cell; Escherichia coli containing the vector of (2); (3) anchoring
CC sequence, or its functional variants or fragments, of at least 8
CC nucleotides that encodes an epitope, a high-affinity binding partner or
CC GST (glutathione-S-transferase); (4) solid phase for immobilizing AS
```

comprising glutathione-Sepharose; (5) combined, stable catalyst (A) comprising SP immobilized on Sepharose for production of poly(1,4-alpha-glucan) (II); and (6) biocatalytic production of (II). AS is used in production of poly(1,4-alpha-glucans) (II), useful for producing films, as food additives, as starting materials for cyclodextrins and as auxiliaries in pharmaceutical formulations. Immobilized AS makes possible efficient, inexpensive and continuous production of poly(1,4-alpha-glucans) (II), and it can be used repeatedly. Compared with known methods, specificity is improved (increased yield of (II) and reduced formation of palatinose) and reaction is complete within 24 hours, compared to 48-72 hours for batch methods. This sequence encodes the expression vector construct pGEX-4T-1-AmSUS which contains the amylosecrase (EC 2.4.1.4) described in the invention.

XX Sequence 6851 BP; 1668 A; 1791 C; 1759 G; 1633 T; 0 other;

# Alignment Scores:

Pred. No.: 0 Length: 6851  
Score: 3360.00 Matches: 626  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.84% Mismatches: 0  
Query Match: 98.59% Indels: 0  
DB: 22 Gaps: 0

US-09-843-007a-2 (1-636) x AAP61712 (1-6851)

QY	10	LeuLeuLeuGlnTyrLeuLysThrArgIleLeuAspIleTyrThrProGluGlnArgAla	29
DB	945	ATGATTTTACAGTACTCAAAACACGATCTTGGACATCTACACGCCCGAACAGCGGCC	1004
QY	30	GlyIleGluLysSerGluAspTyrArgGlnPheSerArgArgMetAspThrHisPhePro	49
DB	1005	GGCATCGAAATCCGAAGACTGGCGGAGTGTTCGCGCGCATGTGATACGCATTTCC	1064
QY	50	LysLeuMetAsnGluLeuAspSerValTyrGlyAsnAsnGluAlaLeuLeuProMetLeu	69
DB	1065	AAACTGATGACGAACTCGACAGGTGTACGACAAACAGACCCCTCTCCCTATGCTG	1124
QY	70	GluMetLeuLeuAlaGlnAlaTyrPdnSerTyrSerGlnArgAsnSerSerLeuLysAsp	89
DB	1125	GAAATGCTGTGCTGGCGCAGGCATGGCAAGCTATTCACCAACGCACTCATCTTAAAGAT	1184
QY	90	IleAspIleAlaArgGluAsnAsnProAspTyrIleLeuSerAsnLysGlnValGlyGly	109
DB	1185	ATCGATATCGCGCGGAAACAAACCCGATTTGGATTTTGTCCAAACAAAGTCGCGGCG	1244
QY	110	ValCysTyrValAspLeuPheAlaGlyAspLeuLysGlyLeuLysAspLysIleProTyr	129
DB	1245	GTGTGCTACGTTGATTTGTTGCGCGGATTTGAAGGCTTGAAGGATATAAATTCCTTAT	1304
QY	130	PheGlnGluLeuGlyLeuThrTyrLeuHisLeuMetProLeuPheLysCysProGluGly	149
DB	1305	TTTCAAGAGCTTGGTTTGACTTATCTGACCTGATGCGCGCTGTGTTAAATGCGCTGAAGGC	1364
QY	150	LysSerAspGlyGlyTyrAlaValSerSerTyrArgAspValAsnProAlaLeuGlyThr	169
DB	1365	AAAGCGACGGCGGTATGCGGTACGAGCTACCGCATGTCAATCGGCACTGGGACACA	1424
QY	170	IleGlyAspLeuArgGluValIleAlaAlaLeuHisGluAlaGlyIleSerAlaValVal	189
DB	1425	ATAGCGACATTGCGCGAAGTCATGCTGCGCTGACGACGCGGCAATTCGCGCTGCTC	1484
QY	190	AspPheIlePheAsnHisThrSerAsnGluHisGluTyrPalaGlnArgCysAlaAlaGly	209
DB	1485	GATTTTATCTTCAACCACTCTCAACGACCAAGATGGGCGCAACGCTGCGCGCGCGG	1544
QY	210	AspProLeuPheAspAsnPheTyrTyrIlePheProAspArgArgMetProAspGlnTyr	229
DB	1545	GACCCGCTTTTCGACAAATTTCTACTATATTTTCCCGACCGCGGATGCGGACCAATAC	1604
QY	230	AspArgThrIleArgGluIlePheProAspGlnHisProGlyGlyPheSerGlnLeuGlu	249

DB	1605	GACCGCACCTTCGCGGAAATCTTCCCGACACGACACCGCGCGGCTTCTCGAACTGGAA	1664
QY	250	AspGlyArgTyrValTyrThrThrPheAsnSerPheGlnTyrAspLeuAsnTyrSerAsn	269
DB	1665	GACGAGCTGGTGTGGACGACCTTCAATTCCTTCCAAATGGGACTTGAATACAGCAAC	1724
QY	270	ProTyrValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAsp	289
DB	1725	CCGTGGGTATTCGCGCAATGGCGGGAATGCTGTCTTCCCACTTGGCGGCTTGC	1784
QY	290	IleLeuArgMetAspAlaValAlaPheIleTyrPlysGlnMetGlyThrSerCysGluAsn	309
DB	1785	ATCTCGGTATGATGCGGTTCCTTTATTGGAAACAAATGGGACCAAGCTCGGAAAC	1844
QY	310	LeuProGlnAlaHisAlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaPro	329
DB	1845	CTGCGCGAGCGCACGCCCTCATCCGCGGTTCATCGCGTATGCGGTATTCGCGCGCC	1904
QY	330	AlaValPhePheLysSerGluAlaIleValHisProAspGlnValValGlnTyrIleGly	349
DB	1905	GCCGTGTTCTCAATCCGAAGCATGTCTCACCCCGACCGAGTCTCTCAATACATCGGG	1964
QY	350	GlnAspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTyrAsnThrIleu	369
DB	1965	CAGACGAATGCCAAATCGGTTCACACCCCTCGCAATGGCATTTGTGGAAACACCTT	2024
QY	370	AlaThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGlu	389
DB	2025	GCCACGCGGAGTCAACCTGTCTCATCGCGCTGACCTACCGCCCAACCTCCCGGAG	2084
QY	390	HisThrAlaTyrValAsnTyrValArgSerHisAspAspIleGlyTyrThrPheAlaAsp	409
DB	2085	CATACCCCTGGGTCAACTACGTTCGCGACGACGACGACGACGACGACGACGACGACG	2144
QY	410	GluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsnArgPhe	429
DB	2145	GAGACCGCGCATATCTGGGCGCTAAGCGGTACGACCGCGCAATTCCTCAACCGCTTC	2204
QY	430	PheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsnProSer	449
DB	2205	TTGCTCAACGTTTCGACGCGCAGCTTCGCTCGTGGGTACCGTTCGAATACACCCCAAGC	2264
QY	450	ThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAlaGlnAspAsp	469
DB	2265	ACAGGCGACGTCGCTGTACGTGTACAGCGCGCGCATTTGGTTCGCGCTTGGCGCAAGACAT	2324
QY	470	ProHisAlaValAspArgIleLysLeuLeuTyrSerIleAlaLeuSerThrGlyGlyLeu	489
DB	2325	CCCCACGCGCTTGACGCGCATCAACCTTGTACAGCATTCGTTGAGTACCGCGGCTCTG	2384
QY	490	ProLeuIleTyrLeuGlyAspGluValGlyThrLeuAsnAspAspAspTyrSerGlnAsp	509
DB	2385	CCGCTGATTTTACCTAGCGCAAGTGGGTACGCTCAATGACGACGACGCTGGTTCGCAAGAC	2444
QY	510	SerAsnLysSerAspAspSerArgTyrAlaHisArgProArgTyrAsnGluAlaLeuTyr	529
DB	2445	AGCAATAAGAGCGACGACGCGCTTGGCGCACCGCTCGCGCTACCAACGAGCCCTGTAC	2504
QY	530	AlaGlnArgAsnAspProSerThrAlaAlaGlyGlnIleTyrGlnGlyLeuArgHisMet	549
DB	2505	GCSCAACGCAACGATCCGTCGACCGCGCGGCAAAATCTATCAGGCGCTTGGCCATATG	2564
QY	550	IleAlaValArgGlnSerAsnProArgPheAspGlyGlyArgLeuValThrPheAsnThr	569
DB	2565	ATTGCGCTCGCCAAAGCAATCCGCGCTTCGACGCGCGGAGGCTGGTTACATTCACACACC	2624
QY	570	AsnAsnLysHisIleIleGlyTyrIleArgAsnAsnAlaLeuLeuAlaPheGlyAsnPhe	589
DB	2625	AACCAACGACACATCATCGCTACATCGCAACATGCGCTTTTGGCATTTCCGCTACTTC	2684
QY	590	SerGluTyrProGlnThrValThrAlaHisThrLeuAlaMetProPheLysAlaHis	609
DB	2685	AGCGAATATCCGCAAAACCGTTTACCGCGCATACCTTCGACGCGCATGCGCTTCACAGGCGCAC	2744



QY 381 LeuThrTyrArgHisAsnLeuProGluHisThrAlaTrpValAsnTyrValArgSerHis 400  
 Db 2070 CTGACCTACCGCCACAACTGCCGAGCATACCGCTGGGTCACTAGCTCCGAGCCAC 2129  
 QY 401 AspAspIleGlyTrpThrPheAlaAspGluAspAlaIleTyrLeuGlyIleSerGlyTyr 420  
 Db 2130 GAGCAGATCGCTGACGCTTGGCCGATGAACCGCGGATATCTGGGATTAAGCGGCTAC 2189  
 QY 421 AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg 440  
 Db 2190 GACCACCGCCAACTCTCAACCGCTTCTTCGTCACCGTTTCGACGGACGCTGCTCGT 2249  
 QY 441 GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla 460  
 Db 2250 GCGGTACCGTTCCATACACCCAGACACAGCGGACCTCCGCTGTGAGTGTACGCGCG 2309  
 QY 461 AlaLeuValGlyLeuAlaGluAspProHisAlaValAspArgIleLeuLeuTyr 480  
 Db 2310 GCATTTGGTGGCTGGCGGACAGACGATCCCGCGCTTGACCGCATCAAACTCTTTGTAC 2369  
 QY 481 SerIleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr 500  
 Db 2370 AGCATTGCTTGGTACCGCGGCTCTGCGCTGATTACCTAGCGACGAAAGTGGGTACG 2429  
 QY 501 LeuAsnAspAspAspTrpSerGlnAspSerAsnIysSerAspAspSerArgTyrAlaHis 520  
 Db 2430 CTCATATGACGACGACTGTGTGCCAAGC-AGCAATAGACGCGACGACGCGTGGGC-CAC 2487  
 QY 521 ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGly 540  
 Db 2488 CGTCGCGCTACACGAGCCCTGTACGGGCAACCGACGATCCGTGACCGCGCGCG 2547  
 QY 541 GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp 560  
 Db 2548 AA-ATCTATCAGGGCTTCGCCATATGATGCGCTCCGCCAAAGCAATCCGCGTTCGAC 2606  
 QY 561 GlyGlyArgGluValThrPheAsnThrAsnAsnIysHisIleIleGlyTyrIleArgAsn 580  
 Db 2607 GCGCGCAGCTGGTGTACATTCACACCAACCAAGCAATCATCGGTACAT-CGCAAC 2665  
 QY 581 AsnAlaLeuAlaPheGlyAsnPheSerGlnTyrProGlnThrValThrAlaHisThr 600  
 Db 2666 AATGCGCTTTTGGCATTCGGTAACCTCAGCGAATATCCGCAACCGTTACCGCGCATACC 2725  
 QY 601 LeuGlnAlaMetProPheIysAlaHisAspIleuIleGlyGlyThrValSerLeuAsn 620  
 Db 2726 CTGCAAGCGCATGCCCTTCAAGCGCAGCACCTCATCGGTGGGCAAACTGTGAGCGCTGAAT 2785  
 QY 621 GlnAspLeuThrLeuGlnProTyrGlnValMetTrpLeuGluIleAla 636  
 Db 2786 CAGGATTTCACGCTTCAGCCCTATCAGGTCTATGCTGCTCGAATCGCC 2833  
 RESULT 8  
 AAT11179  
 ID AAT11179 standard; DNA; 2883 BP.  
 AC AAT11179;  
 XX  
 XX  
 DT 20-JUN-1996 (first entry)  
 DE Neisseria polysaccharea amylosucrase gene.  
 KW Amylosucrase; bacteria; fungi; plants; detection; transformation;  
 KW linear; alpha-1,4-glucans; amylose; sucrose; colourless;  
 KW odourless; tasteless; non-toxic; biodegradable; self-sustaining;  
 KW films; fibres; textiles; paper-making; glass-fibre; tablet binder;  
 KW food thickener; sound proofing; flow properties; paraffin oils;  
 KW organic compound inclusion; chromatographic separation;  
 KW cyclodextrins; ss.  
 XX  
 OS Neisseria polysaccharea.  
 XX  
 XX Key Location/Qualifiers  
 FH

FT CDS 939..2783  
 /\*tag= a  
 XX DB4417879-A1.  
 XX 23-NOV-1995.  
 XX 18-MAY-1994; 94DE-4417879.  
 XX 18-MAY-1994; 94DE-4417879.  
 XX (GENB-) INST GENBIOLOGISCHE FORSCHUNG.  
 XX Buettcher V, Kossmann J, Welsh T;  
 XX P-PSDB; AAR88633.  
 XX WPI: 1996-000447/01.  
 FT New DNA sequence encoding amylo:sucrase of Neisseria - and  
 FT transformed plant, bacteria and fungi able to produce linear  
 FT alpha-1,4-glucan(s), esp. amylose, in practically pure form  
 XX  
 PS Claim 6; Pages 27-33; 42pp; German.  
 XX  
 CC The N. polysaccharea DNA sequence AAT11179, which encodes AAR88633  
 CC amylosucrase (ASA), can be used to produce bacteria, fungi and  
 CC plants that express ASA, and to detect and isolate related DNA  
 CC from other organisms. Transformed plants which express ASA are  
 CC able to produce linear alpha-1,4-glucans, specifically amylose  
 CC from sucrose, which can be used to produce colourless,  
 CC odourless, tasteless, nontoxic, biodegradable, self-sustaining  
 CC films or fibres, e.g. for use in the food, textile, paper-  
 CC making and glass-fibre industries. Amylose can also be used as  
 CC a binder for tablets, thickener for food, in sound proofing  
 CC panels, to improve flow properties in paraffin-based oils, for  
 CC inclusion of organic cpds., in chromatographic sepn. and as a  
 CC starting material for cyclodextrins.  
 XX  
 SQ Sequence 2883 BP; 561 A; 852 C; 758 G; 612 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 0 Length: 2883  
 Score: 3141.50 Matches: 621  
 Percent Similarity: 98.11% Conservative: 3  
 Best Local Similarity: 97.64% Mismatches: 11  
 Query Match: 92.18% Indels: 9  
 DB: 17 Gaps: 1  
 US-09-843-007A-2 (1-636) x AAT11179 (1-2883)  
 QY 1 MetLeuThrProThrGlnGlnValGlyLeuIleLeuGlnTyrLeuIysThrArgIleLeu 20  
 Db 939 ATGTTGACCCCGACGACGAGTCGCTTTGATTTTACGTACTCAAAACAGCATCTTG 998  
 QY 21 AspIleTyrThrProGluGlnArgAlaGlyIleGluIysSerGluAspTrpArgGlnPhe 40  
 Db 999 GACATCTACAGCCCGAAGACGCGCGCGCATCGAAAAATCCGAAGACTGCGCGCAGTTT 1058  
 QY 41 SerArgArgMetAspThrHisPheProIysLeuMetAsnGluLeuAsnSerValTyrGly 60  
 Db 1059 TCGCGCGCATGGATACGCATTTCCCAACTGATGAACGAACTCGACAGGTGTACGCG 1118  
 QY 61 AsnAsnGluAlaLeuLeuProMetLeuGluMetLeuAlaGlnAlaTrpGlnSerTyr 80  
 Db 1119 AACAAAGAGCCCTGCTGCTATCTCGAAATGCTGCTGGCGGAGGATGCGAAAGCTAT 1178  
 QY 81 SerGlnArgAsnSerSerLeuIysAspIleAlaArgGluAsnAsnProAspTrp 100  
 Db 1179 TCCCAAGCAGCACTCATCTTAAAGATATCGATATCGCGCGGAAACAAACCCGATTTG 1238  
 QY 101 IleLeuSerAsnLysGlnValGlyValCysTyrValAspLeuPheAlaGlyAspLeu 120  
 Db 1239 ATTTTGTCCACAAACAAAGTCGCGGGGTGTGCTACGTTGATTTGTTGCCGCGGATTTG 1298

```
QY 121 LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeu 140
Db 1299 AAGGGCTTGAAGAATAAATCTCTATTATTTCAAGAGCTTGGTTGCACTTATCTGCACCTG 1358
QY 141 MetProLeuPheLysCysProGluGlySerAspGlyGlyTyrAlaValSerSerTyr 160
Db 1359 ATGCCGCTGTTTAAATGCCGTGAGGCAAAAGCGACGGGGCTATGCGGTGACGACGTAC 1418
QY 161 ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuA:rgGluValIleAlaAlaLeu 180
Db 1419 CGCGATGTCATTCGGCACTGGGCACAATAGCGACCTTGGCGAAGTCATGCTGCGCTG 1478
QY 181 HisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSerAsnGluHis 200
Db 1479 CACGAATGCG--ATTTCGCGCGGTGCGATTTATCTTCAACCAACCTCCACGACAC 1536
QY 201 GluTTPAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePhe 220
Db 1537 GAATGGCG-CAACGGCTGC--GCCGGCAGCCGGCTTTTCGACAATTTCTACTATATTTTC 1592
QY 221 ProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGln 240
Db 1593 CCCGACCGCGGATCCCGACCAATACGACCCCTCGCGGAATCTTCCCGGACACG 1652
QY 241 HisProGlyGlyPheSerGlnLeuGluAspGlyArgTTPValTTPThrPheAsnSer 260
Db 1653 CACCCGGGGGGCTTCTCGCACTGGAGACGACGCTGGGTGTGGACGACCTTCAATTC 1712
QY 261 PheGlnTTPAspLeuAsnTyrSerAsnProTTPValPheArgAlaMetAlaGlyIleuMet 280
Db 1713 TTCCATGGGACTTGAATTTACAGCAACCGTGGGTATTTCG- GCAATGGCGGGCGGAATG 1771
QY 281 LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTTP 300
Db 1772 CAGTCTCTTCCCACTTGGGCGGTGATCATCTTCGGTATGATGCGGTTCGCTTTATTTGG 1831
QY 301 LysGlnMetGlyThrSerCysGluAsnLeuP:roGlnAlaHisAlaLeuIleArgAlaPhe 320
Db 1832 AAACAATGGGACAGCTGCAAAACCTGCG--GCAGCGACGCGCTCATCCGCGCGTTC 1889
QY 321 AsnAlaValMetArgIleAlaProAlaValPhePheLysSerGluAlaIleValHis 340
Db 1890 AATGCCGTTATGGGTATTCGCGGCGCGCGGTGTTCTTCAATTCGAGCCATGCTCCAC 1949
QY 341 ProAspGlnValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnProLeu 360
Db 1950 CCCGACCAAGTCGTCCAATACATGCGGACGAGCAATGCCAAATCGGTTACAACCCCTG 2009
QY 361 GlnMetAlaLeuLeuTTPAsnThrLeuAlaThrArgGluValAsnLeuHisGlnAla 380
Db 2010 CAAATGGCAATTTGTTGGCAACCTTGCACGCGCGAAGTCAACCTGCTCCATCAGCGG 2069
QY 381 LeuThrTyrArgHisAsnLeuProGluHisThrAlaTTPValAsnTyrValArgSerHis 400
Db 2070 CTGACCTACCGCCACACCTGCGGAGCATACGCGCTGGGTCAACTAGCTCCGCGACCCAC 2129
QY 401 AspAspIleGlyTTPThrPheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyr 420
Db 2130 GACGACATCGGCTGGACGTTTCCGATGAGACGCGCGCATATCTGGGCATAGCGGCTAC 2189
QY 421 AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg 440
Db 2190 GACACCGCAATCTCTCAACCGCTTCTGCTCAACCGTTTCGACGGCAGCTTGGCTGT 2249
QY 441 GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla 460
Db 2250 GCGGTACCGTTCGAATACACACCCAGCAGCGGCGACTGCGGTTCAGTGTACAGCGCGG 2309
QY 461 AlaLeuValGlyLeuAlaGlnAspAspProHisAlaValAspArgIleLysLeuLeuTyr 480
Db 2310 GCATTTGGTCTGGCGCAAGACGATTCCTCCACGCGGTGACCGCATCAAACTCTGTAC 2369
```

```
QY 481 SerIleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr 500
Db 2370 AGCATTCCTTTAGTACCGCGGTCTCGCGCTGATTTACCTAGCGACGAGTGGGTACG 2429
QY 501 LeuAsnAspAspAspTTPSerGlnAspSerAsnLysSerAspAspSerArgTTPAlaHis 520
Db 2430 CTCATGACGACGACTGGTCCCAAGC-AGCAATAAGACGACGACGACCGCTTGGGC-CAC 2487
QY 521 ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGly 540
Db 2488 CGTCCGCGCTACACGAAAGCCCTGTACGCGCAACCCGAAACGATTCGTCGACCGCGCGG 2547
QY 541 GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp 560
Db 2548 AA-ATCTATCAGGGCTTGGCGCATATGATTCGCGTCCGTCGCAAAACAATCCGCGCTTCAC 2606
QY 561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleGlyTyrIleArgAsn 580
Db 2607 GCGGCGAGGCTGGTTACATTCACACCAACCAACGACATCATCGGTACAT-CGCAAC 2665
QY 581 AsnAlaLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThr 600
Db 2666 AATCGGCTTTTGGCATTCGGTAACCTCAGCGAATATCCGAAACCGTTACCGCGCATACC 2725
QY 601 LeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyLysThrValSerLeuAsn 620
Db 2726 CTGCAAGCCCATGCCCTTCAAGGCGACGACCTCATCGGTGGCAAACTGTCAGGCTGAAT 2785
QY 621 GlnAspLeuThrLeuGlnProTyrGlnValMetTTPLeuGluIleAla 636
Db 2786 CAGGATTTGACGCTTCAGCCCTATCAGCTCATGTGGTCGAAATCGCC 2833
RESULT 9
AAS59546/c
ID AAS59546 standard; DNA; 29559 BP.
AC AAS59546;
XX 13-FEB-2002 (first entry)
DT
DE
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant; ds.
OS Propionibacterium acnes.
PN WO200181581-A2.
PD 01-NOV-2001.
PF 20-APR-2001; 2001WO-US12865.
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX (CORI-) CORIXA CORP.
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX WPI; 2001-616774/71.
XX Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX Claim 1; SEQ ID No 41; 1069pp; English.
XX Sequences AAS59506-AAS59804 represent DNA molecules encoding
CC
```



CC Propionibacterium acnes immunogenic polypeptides. The proteins and their  
 CC associated DNA sequences are used in the treatment, prevention and  
 CC diagnosis of medical conditions caused by P. acnes. The disorders include  
 CC SAPHO syndrome (synovitis, acne, pustulosis, hyperostosis and  
 CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved  
 CC in infections of bone, joints and the central nervous system, however it  
 CC is particularly involved in the inflammatory lesions associated with acne  
 CC vulgaris. A method for detecting the presence or absence of P. acnes in a  
 CC patient comprises contacting a sample with a binding agent that binds to  
 CC the proteins of the invention and determining the amount of bound protein  
 CC in the sample. The polypeptides may be used as antigens in the production  
 CC of antibodies specific for P. acnes proteins. These antibodies can be  
 CC used to downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the  
 CC polypeptides shown in AAU49884-AAU50191 and AAU67524-AAU67527.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 29559 BP; 5419 A; 8656 C; 9368 G; 6114 T; 2 other;

# Alignment Scores:

Pred. No.:	3.17e-110	Length:	29559
Score:	1164.00	Matches:	245
Percent Similarity:	57.56%	Conservative:	113
Best Local Similarity:	29.29%	Mismatches:	230
Query Match:	34.15%	Indels:	34
DB:	23	Gaps:	12

US-09-843-007A-2 (1-636) x AAS59546 (1-29559)

Qy	38	ArgGlnPheSerArgArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSer	57
Db	22070	CAGTCTTCGACCTCCGCTGGGAGCACTACTACCTGATCTATGATCGGCTGAGCAGG	22011
Qy	58	ValTyrGlyAsn-----AsnGluAlaLeuLeuProMetLeuGluMetLeuAlaGln	75
Db	22010	GTCTAGCGAGACCGCGCGATGAACCTCGCAGAGATCCGCGCATCTCTCTAAACGG	21951
Qy	76	AlaTrpGlnSerTyrSerGlnArgAsnSerSerLeuLysAspIleAspIleAlaArgGlu	95
Db	21950	TGT-----GCCAGACGCTCTGATGATCTCAAGCGCTCGAGCGCTCGAGACCGCGCTC	21903
Qy	96	AsnAsnProAspTrpIleLeuSerAsnLysGlnValGlyGlyValCysTyrValAspLeu	115
Db	21902	CTGGAGCCCGACTGGCTTCACACACCGACGATGCTGTTATGCACTATACCGACCAT	21843
Qy	116	PheAlaGlyAspLeuLysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeu	135
Db	21842	TTCCTCGGCACCTCAAGGTATAGCGATCACCTCGATCACCTCTGCGACATGGGTGTG	21783
Qy	136	ThrTyrLeuHisLeuSerProLeuPheLysCysProGluGlyLysSerAspGlyTyr	155
Db	21782	CGTATCTGCATCTATCCCTCTGCAACCTCGCCANGTACTGACGATGGTGGCTAT	21723
Qy	156	AlaValSerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGlu	175
Db	21722	GCGGTGGGACCATCGCACTATCGCAGCGATCTGGGTACTACTGATGACCTAGCTGAC	21663
Qy	176	ValIleAlaLeuHisGluAlaGlyIleSerAlaValAlaAspPheIlePheAsnHis	195
Db	21662	CTCAGCGTACCTCGCGGCCACCGCATATCTCTGTCGTCGACCTCATCGTTAACCC	21603
Qy	196	ThrSerAsnGluHisGluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsn	215
Db	21602	GTAGCGCGGACGATGATGGCTCGAGAGCTCGGGCTGTCAACAGAGTACCGCGAC	21543
Qy	216	PheTyrIlePheProAspArgMetProAspGlnTyrAspArgThrLeuArgGlu	235
Db	21542	TATTTCCATATTCTGTGACTCAAGATGAGGTTCAGCGCTCGGAGAAACCTTCGCGAC	21483

Qy	236	IlePheProAspGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArg---TrpVal	254
Db	21482	GTITTTTCGAGCTTCGCGCATGGCAATTCACCTGGGACACGACTGCCAGGTGGGTAT	21423
Qy	255	TrpThrThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheArg	274
Db	21422	TGGCGGACCTTCACAGATTCAGTGGATCTCACTGGGTACCTGACCTGCTTCGC	21363
Qy	275	AlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAsp	294
Db	21362	GAATTCCTGACCTGATCGCTGTCTGGCCATCGCGGGTCGAGGTGTTCGCGCTGCAC	21303
Qy	295	AlaValAlaPheIleTyrLysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHis	314
Db	21302	GCCATCGCTTCATCTGGAAGAACTCGGCACCAACTGTGAGAACCTTCGGAATTCAC	21243
Qy	315	AlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaProAlaValPhePheLys	334
Db	21242	GACATCACTCAGTCATTGCTGAGCAATACGATCGTCGCGCGCTGCTTCATG	21183
Qy	335	SerGluAlaIleValHisProAspGlnValGlnTyrIleGlyGlnAspGlu-----	352
Db	21182	GCGATGCCATCGTTGGCCCGGACGATCTCAGCGGATATTCGAGCGCGGCGCATCGG	21123
Qy	353	-----CysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTrpAsnThrLeu	369
Db	21122	GGAAAGTCTGCGCAGATCTATCAACACAGCTCATGTGTCAGCTGTGGAGCGCCCTA	21063
Qy	370	AlaThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGlu	389
Db	21062	GCTACCGCGACGCTCAGCTCATGGAACACAGCTTGAGTCGGACCGCCGCAACCCCTG	21003
Qy	390	HisThrAlaTrpValAsnTyrValArgSerHisAspAspIleGlyTrpThrPheAlaAsp	409
Db	21002	ACACACACCTGGCGCCACCTAGCTGATGTCACGACGACATCGGATGAGACCGTGTGAT	20943
Qy	410	GluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsnArgPhe	429
Db	20942	GCGACGCGCGTAAACCTGCGCTTGAACCGCTGCGCCACCGCAGGTTCCTCTCGACTTC	20883
Qy	430	PheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsnProSer	449
Db	20882	TATTCGGGACATTCGCCGATCTTCGCCGCTGCTGCTTTCGAGACACACCCGCTT	20823
Qy	450	ThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGly-----Leu	465
Db	20822	ACCGGACCGTTCGATTTAGTGGCTCTTAGCAAGCCTGGCGGGCTGGAATCTGCCTG	20763
Qy	466	AlaGlnAspAspProHisAlaValAsp-----ArgIleLysLeuLeuTyrSer	481
Db	20762	GAGTCCGACACCGACGCTGCTGTGACGCTGCCATCGCCGCGATCGTTATGCTGCACAC	20703
Qy	482	IleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThrLeu	501
Db	20702	GCGNTCTCGGTACGGGAGTACCACTAATCTGGATGGAGACGAGGTTCGGATGCTC	20643
Qy	502	AsnAspAspAspTrpSerGlnAspSerAsnLysSerAspAspSerArgTrpAlaHisArg	521
Db	20642	AAC---GACGACTGGCAACGTCGTCGCGGTCTGCGGACGACACCGCTGGTCCATCGA	20586
Qy	522	ProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsn---AspProSerThrAlaAlaGly	540
Db	20585	CCTATGATGAACCTGGTTCAGTGTCAAGCAGCGCCACCGCGCGCCGACCGTTCCTGCT	20526
Qy	541	GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp	560
Db	20525	CGAATATGGAACGGGTGGCGCGGCGCATCAATGCCGACACCGCGCGCGCTGGTTCAT	20466
Qy	561	GlyGly----ArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArg	579
Db	20465	GCCTCGCTGACACAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	20406
Qy	580	-----AsnAsnAlaLeuLeuAlaPheGlyAsnPheSerGlyTyrProGlnThrVal	596

Db 20405 CCACATCCGAGCGCGCATGATGAGCTATACACATCATGAGCAGCAAGTGTGCTC 20346  
Qy 597 ThrAlaHisThrLeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyLysThr 616  
Db 20345 CCCATGGAACCTTGGCTCT-----GAGCTTGAGGAGCGTCGTCACCGAGTTGCTA 20295  
Qy 617 ValSerLeuAsnGlnAspLeuThr-----LeuGlnProTyrGlnValMet 631  
Db 20294 AGAGGTTGATTCAGACTTGCATCCATCCATGACCTGCGCTCTCCCTATGAGTGCCTT 20235  
Qy 632 TrpLeu 633  
Db 20234 TGGCTA 20229  
RESULT 10  
AA559613  
ID AA559613 standard; DNA; 22934 BP.  
XX AA559613;  
AC AA559613;  
DT 13-FEB-2002 (first entry)  
XX Propionibacterium acnes immunogenic protein encoding DNA #108.  
DE SAPHO syndrome; synovitis; acne; pustulosis; hypervitosis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KW dermatological; osteopathic; neuroprotectant; ds.  
XX Propionibacterium acnes.  
OS WO200181581-A2.  
XX PN 01-NOV-2001.  
PD 20-APR-2001; 2001WO-US12865.  
PF 21-APR-2000; 2000US-199047P.  
PR 02-JUN-2000; 2000US-208841P.  
PR 07-JUL-2000; 2000US-216747P.  
XX (CORI-) CORIXA CORP.  
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX WPI; 2001-616774/71.  
DR Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris -  
XX Claim 1: SEQ ID No 108; 1069pp; English.  
XX Sequences AA559506-AA559804 represent DNA molecules encoding  
CC Propionibacterium acnes immunogenic polypeptides. The proteins and their  
CC associated DNA sequences are used in the treatment, prevention and  
CC diagnosis of medical conditions caused by P. acnes. The disorders include  
CC SAPHO syndrome (synovitis, acne, pustulosis, hypervitosis and  
CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved  
CC in infections of bone, joints and the central nervous system, however it  
CC is particularly involved in the inflammatory lesions associated with acne  
CC vulgaris. A method for detecting the presence or absence of P. acnes in a  
CC patient comprises contacting a sample with a binding agent that binds to  
CC the proteins of the invention and determining the amount of bound protein  
CC in the sample. The polypeptides may be used as antigens in the production  
CC of antibodies specific for P. acnes proteins. These antibodies can be  
CC used to downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the  
CC polypeptides shown in AAU60884-AAU61133 and AAU67652-AAU67654.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 22934 BP; 4591 A; 7273 C; 6479 G; 4591 T; 0 other;  
Alignment Scores:  
Pred. No.: 6,15e-47 Length: 22934  
Score: 560.50 Matches: 162  
Percent Similarity: 43.64% Conservative: 85  
Best Local Similarity: 28.62% Mismatches: 220  
Query Match: 16.45% Indels: 99  
DB: 23 Gaps: 22  
US-09-843-007a-2 (1-636) x AA559613 (1-22934)  
Qy 118 GlyAspLeuIleGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyr 137  
Db 14744 GCGGACTTTAAAGCGCTGACGGGAAACTCGATTAATCTAGTGGCTCGGGGGGATTC 14803  
Qy 138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAVal 157  
Db 14804 CTATGGTTACCACTCTCTATGACTCTCT-----CTCCATGACGAGGTACGTCATT 14857  
Qy 158 SerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIle 177  
Db 14858 CGTGACTATCGCTGGATCCGTGAGGAGCTGGGACCACTTGAAGTTTCCCT 14917  
Qy 178 AlaAlaLeuHisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSer 197  
Db 14918 GAGCAGCCCATGACCGTGGCTTGGCTCATCATCGACTTCGTCATGAACACACCTCG 14977  
Qy 198 AsnGluHisGluTrpAlaGlnArgCysAlaAlaGly-----AspProLeuPheAspAspPhe 216  
Db 14978 GATTCCTCATCGTGGTTCCAGTCTCAGCGCGGATCCGATGCTCTTACGGTAAC--- 15034  
Qy 217 TyrTyrIlePheProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluIle 236  
Db 15035 TACTACGTATGGTCGGAC-----ACTGACGAGCGCTACTCCGATGCTCGATCATC 15085  
Qy 237 PheProAspGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArgTrpValTrp--- 255  
Db 15086 TTCTGGCAT-----ACCGAGGACTCGAATTGGTCTCTGGAT 15121  
Qy 256 ThrThrPheAsnSerPheGlnTrp-----AspLeuAsnTyr 267  
Db 15122 TCCAGCGTAAGCAGATTCTACTGGCACCGGATTTTCCATCCACCGCATCTCACTTT 15181  
Qy 268 SerAsnProTrpValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGly 287  
Db 15182 GAGGAGCTCGAGTTATGAGGAAATGCTCGAGCGCTCGCTTCTGGATGGACTGGGT 15241  
Qy 288 ValAspIleLeuArgMetAspAlaValAlaPheIleTrpLysGlnMetGlyThrSerCys 307  
Db 15242 ATTGACGATTCGAGCTCGAGCGCGCTCCGATCATTCATTGAGCGCGGAGGCGGACGATTC 15301  
Qy 308 GluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPheAsnAlaValMetArgIleAla 327  
Db 15302 GAGAACTTGCAGAACTCACAAGATCTCAAGCAGCTGCGCGCCATGGTCGATGAAGAA 15361  
Qy 328 AlaProAlaValPhePheLysSerGluAlaIleValHisProAspGlnValValGlnTyr 347  
Db 15362 TACCAGCGCGCATCTCTGCTGTGTGAAGCTAACCGAGTGGCTGACGACGCTCGTGAT 15421  
Qy 348 IleGlyGln-----AspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTrp 366  
Db 15422 TTTGGCAACGGTGACGAATGCCAGATGGCTTCCATTTCCGTCATGCCCGCTGTAC 15481  
Qy 367 AsnThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsn 386  
Db 15482 ATGGGACTCGCTCGGTTACGGGAATGCATTAGCAGAGATCTCGGTCACACTCCACAC 15541  
Qy 387 LeuProGluHisThrAlaTrpValAsnTyrValArgSerHisAspAspIleGlyTrpThr 406

Db 15542 ATCCGGAGCGGTCCAGTGGGGGACCTTCTCTGCGCAACACGACGAACTCACCTCGAA 15601  
Qy 407 Phe---AlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPhe 425  
Db 15602 ATGGTGACCGAAGAGATCGTCACTACATGTGG-----15634  
Qy 426 LeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGln 445  
Db 15635 -----GAGGAA 15640  
Qy 446 TyrAsnPro---SerThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGly 464  
Db 15641 TATGCTCTCTGATCCGCTGGGGTGCACATG---GCGATTGCTGCGCGCTGAGCCCT 15697  
Qy 465 LeuAlaGlnAspAspProHisAlaValAspArgIleLeuLeuTyrSerIleAlaLeu 484  
Db 15698 CTGTGCGATACGACGAT-----CGGAAGATCCGGCTACTTAACGCCATGTGTCTG 15748  
Qy 485 SerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThrLeuAsnAspAsp 504  
Db 15749 TCCCTACCGGCTCTCCGGTGTCTGATTACGACGACGAGATTGCG---ATGGGGGACAT 15805  
Qy 505 AspTrpSerGlnAspSerAsn-----LysSerAspAspSerArgTrp 518  
Db 15806 CCTTGGCTGCTGACCGTGCATGGGTGGGTGACCCCGATGCGATGCGATGACTCGGAGACG 15865  
Qy 519 Ala-----519  
Db 15866 GCTGGCTTTTCCATGCGCTTCCAGAGGACTTTCATCTCTACCTCTCATCCGACCTTGGC 15925  
Qy 520 HisArgProArgTyr---AsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThrAla 538  
Db 15926 CACGATCCAGACGACGTCAATGTGCGC-----CGCCAGATGGACGCCGTCTCTCC---15976  
Qy 539 AlaGlyGlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArg 558  
Db 15977 -----TTGCTGGTGTGGAGCGGTGCCATGCTCGGTATCCGCGCCACACCCCGTC 16027  
Qy 559 PheAspGlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIle 578  
Db 16028 TTGGGACCGGCGGACTTCCCGATAGCGGCGCGGACATGGCGGTGATGCTCTCTCTG 16087  
Qy 579 ArgAsnAsn-----AlaLeuLeuAlaPheGlyAsnPheSerGlyTyrProGlnThr 595  
Db 16088 CGTCAACAGCAGCAGAAACGGTCTGCTGCTGGCTAACTTCTCGATACCTGACGCGGATG 16147  
Qy 596 ValThrAlaHisThrLeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyLys 615  
Db 16148 GTTGCCTTTCATCTCCCAATTCGCGGCGCATGACGGGCTCTTCTCTCATCATGTCAG 16207  
Qy 616 ThrValSer---LeuAsnGlnAspLeuThrLeuGln-----ProTyrGlnVal 630  
Db 16208 GACGCGCAACCACTAAAGCTGACGGAACACTGTCTCGTACCGTGTGGGCCATATGGCTAT 16267  
Qy 631 MetTrpLeuGluIleAla 636  
Db 16268 CGATGCGTGCAGAGTGTC 16285

RESULT 11  
ID AAS54127  
ID AAS54127 standard; DNA; 3303 BP.  
AC AAS54127;  
DT 13-FEB-2002 (first entry)  
DE Pseudomonas aeruginosa DNA for cellular proliferation protein #258.  
XX Antisense; ds; prokaryotic cellular proliferation gene;  
KW antibiotic; antibacterial; drug design.  
XX Pseudomonas aeruginosa.

XX W0200170955-A2.  
XX 27-SEP-2001.  
XX 21-MAR-2001; 2001WO-US09180.  
XX 21-MAR-2000; 2000US-191078P.  
XX 23-MAY-2000; 2000US-206848P.  
XX 26-MAY-2000; 2000US-207272P.  
XX 23-OCT-2000; 2000US-242578P.  
XX 27-NOV-2000; 2000US-253625P.  
XX 22-DEC-2000; 2000US-257931P.  
XX 16-FEB-2001; 2001US-269308P.  
XX (ELIT-) ELITEA PHARM INC.  
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
XX Yamamoto RT, Xu HH;  
XX P-PSDB; AAU36268.  
XX WPI; 2001-611495/70.  
XX New polynucleotides for the identification and development of  
XX antibiotics, comprise sequences of antisense nucleic acids -  
XX Claim 27; Seq ID No 7764; 511pp; English.  
XX The invention relates to antisense inhibitors of genes essential to  
XX prokaryotic cellular proliferation, their use in identifying the  
XX genes, their use in the discovery of novel antibiotics, the essential  
XX genes themselves and the encoded proteins. The prokaryotes used are  
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
XX invention is also useful for the identification of potential new targets  
XX for antibiotic development. The antisense nucleic acids can also be used  
XX to identify proteins used in proliferation, to express these proteins,  
XX and to obtain antibodies capable of binding to the expressed proteins.  
XX The programs can be used to screen compounds in rational drug discovery  
XX programmes. The antisense nucleic acid sequence is also useful to screen  
XX for homologous nucleic acids which are required for cell proliferation in  
XX a wide variety of organisms. The present sequence encodes an  
XX essential prokaryotic cellular proliferation protein.  
XX Note: The sequence data for this patent did not form part  
XX of the printed specification, but was obtained in electronic  
XX format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 3303 BP; 578 A; 1171 C; 1084 G; 470 T; 0 other;

Alignment Scores:  
Pred. No.: 3.54e-46 Length: 3303  
Score: 541.50 Matches: 165  
Percent Similarity: 41.60% Conservative: 90  
Best Local Similarity: 26.92% Mismatches: 223  
Query Match: 15.89% Indels: 135  
DB: 23 Gaps: 25  
US-09-843-007A-2 (1-636) x AAS54127 (1-3303)

Qy 96 AsnAsnProAspTrpIleLeuSerAsnLysGlnValGlyValCysTyr-----112  
Db 31 AACGACCGCAATGGTACAGGAC-----GGGTGATCTACAGGTCAC 75  
Qy 113 ValAspLeuPhe-----AlaGlyAspLeuLysGlyLeuLysAsp 125  
Db 76 GGAATCTTCTTACGACCGCCACACAGATGGCATCGGACTTCGGCGGCTCATCGAG 135  
Qy 126 LysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeuMetProLeuPheLys 145  
Db 136 AAGCTCGACTACATCGCGACCTCGCGGTGAACACTCTCTGGCTGCTCGCGTCTAC---192  
Qy 146 CysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyrArgAspValAsnPro 165

Db 193 ---CGTCCGACGCGCGACGCGCTACGACATCGCCGCTACGCTGGCGTGCACAGC 249  
Qy 166 AlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAlaLeuHisGluAlaGlyIle 185  
Db 250 GACTACGCGACGCTCGCGCGCGCGGCTTCATCGCGAGCGCCATCGACGCGGCTG 309  
Qy 186 SerAlaValAlaValAspPheIlePheAsnHisThrSerAsnGluHisGluTrp 202  
Db 310 CGGTGATTACGAGCTGGTATCAACCATCTCCGACCGACATCCCTGGTTATCCGCT 369  
Qy 203 AlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrIlePheProAsp 222  
Db 370 GCCCGCCACGCGAAGAGGATCGCGCGCGCGAC-----TACTAGCTCTGGTTCGAC 423  
Qy 223 ArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGlnHisPro 242  
Db 424 -----ACGACGAGAAATACACAGGTACG---CGATCATCTTCATCGAC----- 465  
Qy 243 GlyGlyPheSerGlnLeuGluAspGlyArgTrpValTrpThr----- 256  
Db 466 -----ACGAGCAGTCCAACTGGACCTGGACCGCGTAGCCCAACAGTAC 510  
Qy 257 -----ThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsnProTrpValPhe 273  
Db 511 TACTGGCACCCTTCTATTCCACGACGCGGACCTGAACCTTCGACAAACCGCAGGTCTG 570  
Qy 274 ArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMet 293  
Db 571 CGCAGGTGCTCGGGTGTGCTACTGCTGCTGACATGGGGTGCAGCGCTCGCGCTG 630  
Qy 294 AspAlaValAlaPheIleTrpIysGlnMetGlyThrSerCysGluAsnLeuProGlnAla 313  
Db 631 GACGGGATTCCTACTGATCGAAGCGACGCGACCGACGACGAGCAACCTCGCGAGACC 690  
Qy 314 HisAlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaProAlaValPhe 333  
Db 691 CACCAAGGTCTCAAGCGCATCCGCGCGAGCTGGACCGCGCTACTTCGCGCGGAGATGCG 750  
Qy 334 LysSerGluAlaIleValHisProAspGlnValValGlnTyrIleGly----- 349  
Db 751 CTGGCGAGGCAATCAGTGGCGGAGACAGACCGCGCGTACTTCGCGCGGAGATGCG 810  
Qy 350 -----GlnAspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuTrpAsn 367  
Db 811 GCGAGGGCGAGAAATGCGACATGGCTTCCACTTCCTCCGCTGATCGCGCATGTACATG 870  
Qy 368 ThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeu 387  
Db 871 GCATCGCCACGAGGATCGCTATCCGATCCGACATCTCGCGCAGACCCGCGACATC 930  
Qy 388 ProGluHisThrAlaTrpValAsnTyrValArgSerHisAspIleGlyTrpThrPhe 407  
Db 931 CGGGCAATTGCCAATGGCGGATCTTCCTGCGCAACCCACGACGAGCTGACCCCTGGAGATG 990  
Qy 408 AlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsn 427  
Db 991 GTCACCGAC-----GACGAGCGGACATCTC----- 1017  
Qy 428 ArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsn 447  
Db 1018 -----TGGAAC----- 1023  
Qy 448 ProSerThrGlyAspCysArgValSer-----GlyThrAlaAlaAlaLeuValGly 464  
Db 1024 CACTATGCGCGCGACGCGCGCGCGCTCAACCTGGGATCGCGCGCGCTGGCGCG 1083  
Qy 465 LeuAlaGlnAspAspProHisAlaValAspArgIleLysLeuLeuTyrSerIleAlaLeu 484  
Db 1084 CTGGTGGAGGTGACCGCGCG-----CGCATCGAGCTGTCACAGCTGCTGCTGTG 1134  
Qy 485 SerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr----- 500

Db 1135 TCGATCGCGGCGACGCGACCTGTACTACGCGCGAGAGATCGGCATCGCGACACATC 1194  
Qy 501 ---LeuAsnAspAspAsp-----TrpSerGlnAspSerAsn--- 511  
Db 1195 TACTCTCGCGACCGCGACGCGCGCGGACCCCGATCGATCGTGGTGGACCGCACGCG 1254  
Qy 512 -----LysSerAspAspSerArgTrpAlaHisArgProArgTyrAsnGluAlaLeu 528  
Db 1255 GGCTTCTCCGCGCGCGACCGCGCGAGCTGGTG---CTGCCGCGGATTCGACCCCGCTG 1311  
Qy 529 Tyr-----AlaGlnArgAsnAspProSerThrAlaAlaGly 540  
Db 1312 TACGGCTACACGACGATCAACGTCGAGCGCGACGCGCGCGCGCATTCG----- 1362  
Qy 541 GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp 560  
Db 1363 ---CTGCTCACTGATCGCGCGCTCTCGCGTGGCGACCGACGAGAGCGCTTCGCG 1419  
Qy 561 GlyGlyArgLeuValThrPheAsnThrAsnLysHisIleIleGlyTyrIle----- 578  
Db 1420 CGCGGACCTGAGATGCTCGCGCGCGACGCGACCGCATCTCGCTACCTGCGCGAA 1479  
Qy 579 -----ArgAsnAsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrPro 593  
Db 1480 TACCGCGCGCGACGCGAGGACGATCTCTGCTGGCCCACTGTGCGCGCGCGCC 1539  
Qy 594 GlnThrValThrAlaHisThrLeuGlnAlaMetProPheLysAlaHis----- 609  
Db 1540 CAGCGCGTG-----GAACTGGACCTCGCGACGCACTGCGCGCAAGGTG 1581  
Qy 610 -----AspLeuIleGlyGlyLysThrVal-----SerLeuAsnGlnAspLeu 623  
Db 1582 CCGGTGAGATGATCGCGCGCATGCTTCCCGCGATCGCGGAGCTGACTACTCTGCTG 1641  
Qy 624 ThrLeuGlnProTyrGlnValMetTrpLeuGluIleAla 636  
Db 1642 ACCCTCGCGCTTCTACTGTTCTTACTCGCGCC 1680

RESULT 12  
AAH67492  
ID AAH67492 standard; DNA; 1794 BP.  
XX  
AC AAH67492;  
XX  
DT 26-SEP-2001 (first entry)  
XX  
DE C glutamicum coding sequence fragment SEQ ID NO: 2527.  
XX  
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
KW organic acid synthesis; ds.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN EP1108790-A2.  
XX  
PD 20-JUN-2001.  
XX  
PP 18-DEC-2000; 2000EP-0127688.  
XX  
PR 16-DEC-1999; 99JP-0377484.  
PR 07-APR-2000; 2000JP-0159162.  
PR 03-AUG-2000; 2000JP-0280988.  
XX  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
XX  
XX WPI; 2001-376931/40.  
DR P-PSDB; AAG92273.  
XX  
PT Novel polynucleotides derived from Coryneform bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing

expression profile or pattern of a gene and identifying homologous gene

Claim 8, SEQ ID NO: 2527; 246pp + Sequence Listing; English.

The present invention provides a number of nucleotide and protein sequences from the Corynebacterium Corynebacterium glutamicum. These sequences are useful for identifying the mutation point of a gene derived from a mutant of corynebacterium, measuring expression amount and analysing the expression profile or expression pattern of a gene derived from Corynebacterium bacterium, and identifying a homolog of a gene derived from corynebacterium bacterium. Corynebacterium bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a nucleic acid described in the exemplification of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office.

Sequence 1794 BP; 425 A; 542 C; 432 G; 395 T; 0 other;

#### Alignment Scores:

Pred. No.:	2.64e-45	Length:	1794
Score:	529.50	Matches:	153
Percent Similarity:	46.32%	Conservative:	105
Best Local Similarity:	27.47%	Mismatches:	218
Query Match:	15.54%	Indels:	81
DB:	22	Gaps:	22

US-09-843-007a-2 (1-636) x AAH67492 (1-1794)

QY	118	GlyAspLeuLysGlyLeuLysAspLeuLysPheGlnGluLeuGlyLeuThrTyr	137
DB	223	GGATCGTTGAAGGCTGACCGAAGAACTGGATTACATCCAGTGGCTCGCGTGATTC	282
QY	138	LeuHisLeuMetProLeuPheLysCysProGluLysSerPheGlyTyrAlaVal	157
DB	283	ATTGGATCCCACTGTTTATGATCCCTCCCA-----CTGCGCGGACGCGGTACGATC	336
QY	158	SerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValile	177
DB	337	CGCAACTTCCTGTAATCTGCGCGAATTCGGACCGTCGATGACTTCGCGGAATCGTT	396
QY	178	AlaAlaLeuHisGluAlaGlyIleSerAlaValAspPheLeuPheAsnHisThrSer	197
DB	397	GACACGCCACCGCGCTGCGTGTATATCACCGACTTGTGTCATGATCATCACCTCC	456
QY	198	AsnGluHisGluTyrAlaGlnArgCysAlaAlaGlyAspProLeu-----PheAspAsn	215
DB	457	GACCAGACGATGTTCCAGAA---TCCCGCGCGACCCACCGGCCCTTACGGAGAT	513
QY	216	PheTyrThrPheProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGlu	235
DB	514	TTCTATGTGTGGAGCGATGAT-----CCACCCCTGTACAAACGAGCC---CGCATC	561
QY	236	IlePheProAspGlnHisProGlyGlyPheSer---GlnLeuGluAspGlyArgTyrVal	254
DB	562	ATCTTGTAGATACAGAAAGATCCACTGACCTATGATCCGGTGGTGGCCAGTACTTC	621
QY	255	TyrThrThrPheAsnSerPheGlnTyrAspLeuAsnTyrSerAsnProTyrValPheArg	274
DB	622	TGGCACCCTTCTCTCCCAACCAAGACCTCACTACGACCAACCCGCGAGTCCAGAG	681
QY	275	AlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAsp	294
DB	682	GCCATGCTAGATGCTTCGCTTCTGGCTGACCTGGGACTTGTATGGTTTCCGACTAGAT	741
QY	295	AlaValAlaPheIleTyrLysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHis	314
DB	742	GCCTTCCTTATCTTTTGAACGGAGGACCAACGCGGAAACCTCAAGAAACCCAC	801
QY	315	AlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaValProAlaValPhePheLys	334

RESULT 13  
ABS65343  
ID ABS65343 standard; DNA; 1981 BP.

DB	802	GATTTCCTCAAACTGTGTGCTCTGTCTATTGAGAGGATATACCCGCGCGATCTGCTC	861
QY	335	SerGluAlaIleValHisProAspGlnValValGlnTyrIleGlyGln-----	350
DB	862	GCAGAACCAACCAATGCGCCCAAGATGTGCTCGAATACTTCGTTGAAAAAGACAAAGC	921
QY	351	AspGluCysGlnIleGlyTyrAsnProLeuGluMetAlaLeuLeuTyrAsnThrLeuAla	370
DB	922	GATGATGCGCATGGCTTCCACTTCCCTTTCGATGCGCGGATCTTCATGGAGTTCG	981
QY	371	ThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHis	390
DB	982	CAAGGTTCAAGCACCCTGATGAGTCTGCGCAACACCCCGGAGATTCACCAAGACT	1041
QY	391	ThrAlaTyrValAsnTyrValArgSerHisAspAspIleGlyTyrThrPhe---AlaAsp	409
DB	1042	GCCCAATGGGTATTTCTCGTATCATGATGAGTCACTCCCTTGAATGGTCTCGGAT	1101
QY	410	GluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsn-----	427
DB	1102	GAGGAACGCGATACATG-----TACTCCCAATTGCGCTCCGAACCT	1143
QY	428	ArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsn	447
DB	1144	CGCATCGCGCCCAAC-----	1158
QY	448	ProSerThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAlaGln	467
DB	1159	-----GTAGGAATCCGCGAGCGCTTTCCTCCACTGCTGAA	1194
QY	468	AspAspProHisAlaValAspArgIleLysLeuLeuTyrSerIleAlaLeuSerThrGly	487
DB	1195	GGCGACCGC-----AACCACTGGAAGTCTTCAAGGTTTGTGTCTCTACCT	1245
QY	488	GlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr-----LeuAsn	502
DB	1246	GGCTCACCGGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1305
QY	503	AspAspAsp-----TyrSerGlnAspSerAsn-----	511
DB	1306	GACCGCGACGAGTGGCGACCCCATCGATGCTGCTCAACGACCGCAACGGTGTCTCC	1365
QY	512	LysSerAspAspSerArgTyrAlaHisArgProArgTyrAsnGluAlaLeu---TyrAla	530
DB	1366	AAAGCTGATCTTGAAGCGCTTACCTTCCAGCGATCCAAATGATCAATACGGCTACGC	1425
QY	531	GlnArgAsnAspProSer-----ThrAlaAlaGlyGlnIleTyrGlnGlyLeuArgHis	548
DB	1426	CAAGTAAACGTGGAAGCAACTCAACCGCGAAACTCCCTGCTGGCTGGCTCCGAAC	1485
QY	549	MetIleAlaValArgGlnSerAsnProArgPheAspGlyGlyArgLeuValThrPheAsn	568
DB	1486	CAATCTTATCCGCAAGCAGTACCGCGCATTTGGTGGCGGAACCTACCGTGAAGTCTC	1545
QY	569	ThrAsnAsnLysHisIleIleGlyTyrIleArgAsnAsn-----AlaLeuLeuAla	585
DB	1546	TCACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1605
QY	586	PheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThrLeuGlnAlaMetPro	605
DB	1606	GTCACCAACATGACCAATATCTCAGGCGATC-----TCGCTTATGTTGCTGAA	1656
QY	606	PheLysAlaHis-----AspLeuIleGlyGlyLys-----ThrValSerLeu	619
DB	1657	TTTGACGAGCACACCCCTCGAGAGATGTCGGCGGCGAGCTTTCCTACCATTTGCTGAA	1716
QY	620	AsnGln---AspLeuThrLeuGlnProTyrGlnValMetTyrLeuGluIle	635
DB	1717	CGGAGTGGATGTCTACCTTATGACCTTACCGGATTTCTCTGTTGATCTC	1767

XX AC ABS65343;  
 XX DT 15-NOV-2002 (first entry)  
 XX DE DNA encoding C. glutamicum metabolic pathway (MP) protein #2.  
 XX KW Metabolic pathway protein; MP; cell metabolism; amino acid; vitamin;  
 XX KW cofactor; nucleoside; nucleoside; trehalose; fine chemical production;  
 XX KW organic acid; non-proteinogenic amino acid; purine base; carbohydrate;  
 XX KW pyrimidine base; lipid; unsaturated fatty acid; diols; polyketide;  
 XX KW aromatic compound; food industry; animal feed; cosmetic industry;  
 XX KW pharmaceutical industry; gene; ds.  
 XX OS Corynebacterium glutamicum ATCC 13032.  
 XX PN W0200251231-A1.  
 XX PD 04-JUL-2002.  
 XX PF 22-DEC-2000; 2000WO-EPI3143.  
 XX PR 22-DEC-2000; 2000WO-EPI3143.  
 XX PA (BADI ) BASF AG.  
 XX PI Pompejus M, Kroeger B, Zeider O, Schroeder H;  
 XX DR WPI; 2002-643289/69.  
 XX DR P-PSDB; ABS80322.  
 XX PT New metabolic pathway genes of Corynebacterium glutamicum for producing  
 XX PT fine chemicals, e.g. lipids, (un)saturated fatty acids, vitamins,  
 XX PT cofactors or enzymes used in food, feed, cosmetics or pharmaceutical  
 XX PT industries -  
 XX PS Claim 1; Page 95-98; 176pp; English.  
 XX CC The present invention relates to the isolation of Corynebacterium  
 XX CC glutamicum metabolic pathway (MP) proteins, and the polynucleotide  
 XX CC sequences encoding them. The MP proteins are enzymes involved in  
 XX CC the metabolism of molecules important for the normal functioning  
 XX CC of cells (e.g. amino acids, vitamins, cofactors, nucleosides and  
 XX CC nucleosides, or trehalose). The polynucleotide sequences encoding  
 XX CC the MP proteins are useful for producing fine chemicals, particularly  
 XX CC organic acids, non-proteinogenic amino acids, purine and pyrimidine  
 XX CC bases, nucleosides, nucleotides, lipids, (un)saturated fatty acids,  
 XX CC diols, carbohydrates, aromatic compounds, vitamins, cofactors,  
 XX CC polyketides and enzymes. The fine chemicals are useful in the food,  
 XX CC animal feed, cosmetic or pharmaceutical industries. ABS65342-ABS65364  
 XX CC encode the C. glutamicum MP proteins of the invention.  
 XX SQ Sequence 1981 BP; 470 A; 596 C; 476 G; 439 T; 0 other;

Alignment Scores:  
 Pred. No.: 3,05e-45 Length: 1981  
 Score: 529.50 Matches: 153  
 Percent Similarity: 46.32% Conservative: 105  
 Best Local Similarity: 27.47% Mismatches: 218  
 Query Match: 15.54% Indels: 81  
 DB: 24 Gaps: 22

US-09-843-007A-2 (1-636) x ABS65343 (1-1981)

QY 118 GlyAspLeuLysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyr 137  
 DB 380 GGATCGTTGAAGGCGCTGACCGGAAACACTGGATTACATCCAGTGGCTCGCGTGGATTGC 439  
 QY 138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaVal 157  
 DB 440 ATTGGATCCACCGCTTTTATGATTCCCA-----CTGCGCGACGCGGTTACGATATC 493  
 QY 158 SerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIle 177

DB 494 CGCAACTTCGGTGAATCCTGCGCAATTCGGCACCGTCGATGACTTCGTGGAATCGTT 553  
 QY 178 AlaAlaLeuHisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSer 197  
 DB 554 GACCAAGCCACCGCGCTGCTGCTGTTATCACCGACTTGGTCATGAATCACACTCC 613  
 QY 198 AsnGluHisGluTTPAlaGlnArgCysAlaAlaGlyAspProLeu-----PheAspAsn 215  
 DB 614 GACCAAGCCACCGCTGCTGCTGTTATCACCGACTTGGTCATGAATCACACTCC 670  
 QY 216 PheTyrTyrIlePheProAspArgMetProAspGlnTyrAspArgThrLeuArgGlu 235  
 DB 671 TTTATGTGTGGAGCGATGAT-----CCACCCCTGTACACGAGGCC---CGCATC 718  
 QY 236 IlePheProAspGlnHisProGlyGlyPheSer---GlnLeuGluAspGlyArgTyrVal 254  
 DB 719 ATCTTTGTAGATACAGAGAATCCAACTGGACCTATGATCCGGTCGGTGGCCAGTACTTC 778  
 QY 255 TrpThrThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsnProTyrValPheArg 274  
 DB 779 TGGCACCGCTTCTCTCCACCCACGACCTCACTACGACACACCCCGCATCCAGAG 838  
 QY 275 AlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAsp 294  
 DB 839 GCCATGCTAGATGCTCTGCGTTCTGGCTGGACCTGGAGCTTGATGGTTTCGAGTAGAT 898  
 QY 295 AlaValAlaPheIleTyrLysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHis 314  
 DB 899 GCGGTTCCTTATCTTTTGAACGCGAGCCACCAACGCGGAAACCTCAAGAAACCCAC 958  
 QY 315 AlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaProAlaValPhePheLys 334  
 DB 959 GATTTCTCCAACTGTGCTGCTCTCATTTGAGAGGAATACCCCGCGCATCTCTGCTC 1018  
 QY 335 SerGluAlaIleValHisProAspGlnValGlnTyrIleGlyGln----- 350  
 DB 1019 GCAGAGCGCAACCAATGCGCCCAAGATGTGTGCGAATACTTCGTGAAACAGCAAGGC 1078  
 QY 351 AspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTyrAsnThrLeuAla 170  
 DB 1079 GATGAATGCCATGCGCTTCCACTTCCCTTTGATGCGCGCATCTTCCATGGAGATTGCG 1138  
 QY 371 ThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHis 390  
 DB 1139 CAAGTTCCAGCACCCCGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1198  
 QY 391 ThrAlaTrpValAsnTyrValArgSerHisAspAspIleGlyTyrThrPhe---AlaAsp 409  
 DB 1199 GCGCAATGGGTATTTCTGCGTATCATGATGATGATGATGATGATGATGATGATGATGAT 1258  
 QY 410 GluAspAlaAlaTyrLeuGlyLysSerGlyTyrAspHisArgGlnPheLeuAsn----- 427  
 DB 1259 GAGGAACGCGCTACATG-----TACTCCCAATTCGCTCCGAACT 1300  
 QY 428 ArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsn 447  
 DB 1301 CGCATGCGCGCCAC----- 1315  
 QY 448 ProSerThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAlaGln 467  
 DB 1316 -----CTAGCAATCCGAGCGCTTTTCCCACTGCTTGA 1351  
 QY 468 AspAspProHisAlaValAspArgIleLysLeuTyrSerIleAlaLeuSerThrGly 487  
 DB 1352 GCGCAGCCG-----AACCATGGAACCTCTTACGGTTTGTGCTGCTCTACTCT 1402  
 QY 488 GlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr-----LeuAsn 502  
 DB 1403 GGCCTACCGTGTGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1462  
 QY 503 AspAspAsp-----TrpSerGlnAspSerAsn----- 511





```
QY 371 ThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHis 390
Db 339094 CRAAGTTCACGCCCGCATCAGTATGATCTCGGCCAACACCCCGAGATTCACAGACT 339153
QY 391 ThrAlaTTPValAsnTyrValArgSerHisAspAlleGlyTTPhrPhe---AlaAsp 409
Db 339154 GCCCAATGGGTATTTCTCGGTATATCATGATGAGCTCACCTTGAATGGTCTCCGAT 339213
QY 410 GluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsn---- 427
Db 339214 GAGGAACGACGCTACATG-----TATCCCAATTCGCTCCGAACCT 339255
QY 428 ArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsn 447
Db 339256 CGCATCGCGCCCAAC----- 339270
QY 448 ProSerThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAlaGln 467
Db 339271 -----GTAGGAATCCGCGAGCGCTTTCGCCACTGTTGAA 339306
QY 468 AspAspProHisAlaValAspArgIleLysLeuLeuTyrSerIleAlaLeuSerThrGly 487
Db 339307 GCGCAGCGC-----AACCACTGGAATCTCTCACGGTTGTGTCTCTACT 339357
QY 488 GlyLeuProLeuLeuTyrLeuGlyAspGluValGlyThr-----LeuAsn 502
Db 339358 GGTCAACCGTGTGTATTACGGTGAATGAATGCGATCGGCACAAATATCGGTCTCCAC 339417
QY 503 AspAspAsp-----TTPSerGlnAspSerAsn----- 511
Db 339418 GACCGCAGGAGTGCACCCCGCATGTCAGTGTCTCAACGACGCGCAACGGTGTCTCTCC 339477
QY 512 LysSerAspAspSerArgTTPAlaHisArgProArgTyrAsnGluAlaLeu---TyrAla 530
Db 339478 AAAGCTGATCTTGAGCGCTGTACCTTCCAGCGATCCAAATGATCAATACGCGTACGCC 339537
QY 531 GlnArgAsnAspProSer-----ThrAlaAlaGlyGlnIleTyrGlnGlyLeuArgHis 548
Db 339538 CAGTAAACCTGGAAGACCACTCAACCGGAAACTCCCTGCTGCGCTGCTCCGAAC 339597
QY 549 MetIleAlaValArgGlnSerAsnProArgPheAspGlyGlyArgLeuValThrPheAsn 568
Db 339598 CAAATCTTATCGCAACGAGTACCGCGCATTTGTGCGGAACTACCGTGAAGTCTCC 339657
QY 569 ThrAsnAsnLysHisIleIleGlyTyrIleArgAsnAsn-----AlaLeuLeuAla 585
Db 339658 TCACCAATGAGTCAGTGTGACATTTTACGAGAAACACAAAGGCGCAACCACTTTGTGT 339717
QY 586 PheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThrLeuGlnAlaMetPro 605
Db 339718 GTCAACACATGAGCAATATCTCAGCGATC-----TCGCTTGTATTGCTGTAA 339768
QY 606 PheLysAlaHis-----AspLeuIleGlyGlyLys-----ThrValSerLeu 619
Db 339769 TTTGCAAGGACACACCCCTCGAGAGATGTCGGCGCGGCGAGTGTCTCCCTACCATGCTGAA 339828
QY 620 AsnGln---AspLeuThrLeuGlnProTyrGlnValMetTyrLeuGluIle 635
Db 339829 CGGAGTGGATGTCATTTAGCCCTCAGCGATTCCTTGTGTTGATCTC 339879
RESULT 15
AAH68532
ID AAH68532 standard; DNA; 349980 BP.
XX
AC AAH68532;
XX
XX 26-SEP-2001 (first entry)
DE C glutamicum coding sequence fragment SEQ ID NO: 7067.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
```

```
XX Corynebacterium glutamicum.
OS
XX
XX EP1108790-A2.
XX
XX 20-JUN-2001.
XX
XX 18-DEC-2000; 2000EP-0127688.
XX
XX 16-DEC-1999; 99JP-0377484.
XX
XX 07-APR-2000; 2000JP-0159162.
XX
XX 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW ) KYOWA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX WPI; 2001-376931/40.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
XX mutation point of a gene, measuring expression of a gene, analysing
XX expression profile or pattern of a gene and identifying homologous gene
XX
XX
XX Disclosure; SEQ ID NO: 7067; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
XX sequences from the Coryneform bacterium Corynebacterium glutamicum. These
XX are useful for identifying the mutation point of a gene derived from a
XX mutant of coryneform bacterium, measuring expression amount and
XX analysing the expression profile or expression pattern of a gene derived
XX from Coryneform bacterium, and identifying a homologue of a gene derived
XX from coryneform bacterium. Coryneform bacteria are useful for producing
XX amino acids, nucleic acids, vitamins, saccharides and organic acids,
XX particularly L-lysine. The present sequence is a nucleic acid described
XX in the exemplification of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX European Patent Office.
XX
XX SQ Sequence 349980 BP; 80900 A; 98397 C; 92139 G; 78544 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 6,14e-42 Length: 349980
XX Score: 529.50 Matches: 153
XX Percent Similarity: 45.32% Conservative: 105
XX Best Local Similarity: 27.47% Mismatches: 218
XX Query Match: 15.54% Indels: 81
XX DB: 22 Gaps: 22
XX
XX US-09-843-007A-2 (1-636) x AAH68532 (1-349980)
QY 118 GlyAspLeuLysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyr 137
Db 38335 GGATCGTTGAAGGCTGACCGAATACTGATACATCCAGTGGCTCGGGGTGATTCG 38394
QY 138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAlaVal 157
Db 38395 ATTTGGATCCACCGCTTTTATGATTCCTCA-----CTGCGCAGCGGGGTACCATATC 38448
QY 158 SerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIle 177
Db 38449 CGCAACTTCGGTGAATCTCGCCGAATTCGCGACCGTCGATGATCTTGTGGAATCGTT 38508
QY 178 AlaAlaLeuHisGluAlaGlyIleSerAlaValAspPheIlePheAsnHisThrSer 197
Db 38509 GACCACGCCACCGCGGTGGCTGTGTTATCATCCGACTTGTGTATGATATCACACCTCC 38568
QY 198 AsnGluHisGluTTPAlaGlnArgCysAlaAlaGlyAspProLeu-----PheAspAsn 215
Db 38569 GACCACGACGATGTTTCCAAAGAA---TCCGCGCGCGACCCCAACCGGCCCTTACGGAGAT 38625
```

```
QY 215 PheTyrTyrIlePheProAspArgMetProAspGlnTyrAspArgThrLeuArgGlu 235
DB 38626 TTCTATGTGTGGAGCGATGAT-----CCACCGCTGTACAGAAAGCC---GCGATC 38673
QY 236 IlePheProAspGlnHisProGlyGlyPheSer---GlnLeuGluAspGlyArgTrpVal 254
DB 38674 ATCTTTGTAGATACAGAAAGTCCAACTGGACCTATGATCGGTGCGCGATCTTC 38733
QY 255 TrpThrThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheArg 274
DB 38734 TGGCACCGCTTCTCTCCCAACAGACAGACCTCAACTAGACACACCGCGAGTCCAAAG 38793
QY 275 AlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAsp 294
DB 38794 GCGATGTAGATGCTTGGCTTCTGGCTGGACCTGGACCTGATGGTTCCGATAGAT 38853
QY 295 AlaValAlaPheIleTrpHisGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHis 314
DB 38854 GCGGTTCTCTATCTTTTGAACGCGAAGGCACCAACGCGGAAACCTCAAGAAACCCAC 38913
QY 315 AlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaProAlaValPhePheLys 334
DB 38914 GATTTCCCAACTGTGTGCTCTGTGATAGAGAGGAATACCCGCGCGAATCTGCTC 38973
QY 335 SerGluAlaIleValHisProAspGlnValValGlnTyrIleGlyGln----- 350
DB 38974 GCAGAAGCCACCAATGCCCCAGATGTGTGATATCTCGGTGAAAAGACACAAAGCC 39033
QY 351 AspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTrpAsnThrLeuAla 370
DB 39034 GATGAATCCACATGCGCTTCCACTTCCCTTTGATGCGCGCGATCTTTCATGGAGTTCGC 39093
QY 371 ThrArgGluValAsnLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHis 390
DB 39094 CNAAGTTACGCGACCCCGATAGTGGATCTTGGCCACACCCCGAGATTTCCCAAGACT 39153
QY 391 ThrAlaTrpValAsnTyrValArgSerHisAspIleGlyTrpThrPhe---AlaAsp 409
DB 39154 GCGCAATGGGTATTTCTCGGTAAATCATGATGAGCTCACCCTTGAAATGGTCTCCGAT 39213
QY 410 GluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsn----- 427
DB 39214 GAGGAACGAGCTACATG-----TACTCCCAATTGCTTCCGCAACT 39255
QY 428 ArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsn 447
DB 39256 CGCATGCGCGCCAAC----- 39270
QY 448 ProSerThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAlaGln 467
DB 39271 -----GTAGGAATCCGACGCGCTTTCCCCACTGCTTGAA 39306
QY 468 AspAspProHisAlaValAspArgIleLysLeuLeuTyrSerIleAlaLeuSerThrGly 487
DB 39307 GCGGACCGC-----AACGAGTGAACCTCTTCAAGGTTGTTGCTCTCTACT 39357
QY 488 GlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr-----LeuAsn 502
DB 39358 GCGTCACCGGTGTTGTTATTACGGTATGAATGGCATGGCGGACAAATATCTGGCTCCAC 39417
QY 503 AspAspAsp-----TrpSerGlnAspSerAsn----- 511
DB 39418 GACCGGACGGAGTGGCGACCCCATGAGTGGTCCACGACCCGACCGGTTGTTCTCC 39477
QY 512 LysSerAspAspSerArgTrpAlaHisArgProArgTyrAsnGluAlaLeu---TyrAla 530
DB 39478 AAAGCTGATCTGAACGCGTGTACCTCCAGCGATCCAAATGATCAATACGGGTACGCC 39537
QY 531 GlnArgAsnAspProSer-----ThrAlaAlaGlyGlnIleTyrGlnGlyLeuArgHis 548
DB 39538 CAAGTAAACGTTGGAAGGCAACTCAACCGCAAAACTCCCTGCTGCTGCTCGGAAC 39597
QY 549 MetIleAlaValArgGlnSerAsnProArgPheAspGlyGlyArgLeuValThrPheAsn 568
```

```
DB 39598 CAAATCCTTATCCGCAAGCAGTAGTACCGCGCATTTTGGTCCGGAACCTACCGTGAAGTGTCC 39657
QY 569 ThrAsnAsnLysHisIleIleGlyTyrIleArgAsnAsn-----AlaLeuLeuAla 585
DB 39658 TCCACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 39717
QY 586 PheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThrLeuGlnAlaMetPro 605
DB 39718 GTCAACAACATGAGCAAAATATCTCTCAGCGAGTC-----TCGCTTGTGATTGGTGAA 39768
QY 606 PheLysAlaHis-----AspLeuIleGlyGlyLys-----ThrValSerLeu 619
DB 39769 TTTCCAGACACACCCCTCGAGAGATGTCGGGCGGAGCTGTTCCCTACCATTTGCTGAA 39828
QY 620 AsnGln---AspLeuThrLeuGlnProTyrGlnValMetTrpLeuGluIle 635
DB 39829 CCGGAGTGGATTGTCACTTTTAGCCCTCAGCGATTCTTCTGTTGATCTC 39879
```

Search completed: November 8, 2003, 20:28:51  
Job time : 852 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 8, 2003, 19:59:38 ; Search time 110 Seconds  
(without alignments)  
2551.997 Million cell updates/sec

Title: US-09-843-007A-2

Perfect score: 3408

Sequence: 1 MLTPTQVGLILQVLTIRL.....VSLNQDLTLPQVYVWLEIA 636

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139986

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-D=/cgn2\_1/USPTO.spool.p/US09843007/runat\_07112003\_170420\_25382/app\_query.fasta\_1.775  
-DB=Issued Patents NA -QFWT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0  
-LIST=45 -DALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09843007 @CEN 1 1 103 @runat\_07112003\_170420\_25382 -NCPU=6 -ICPU=3  
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3408	100.0	2914	3	US-08-737-752A-1
2	543.5	15.9	3414	4	US-09-252-991A-8258
3	543.5	15.9	4125	4	US-09-252-991A-8065
4	509.5	15.0	4403765	3	US-09-103-840A-2
5	509.5	15.0	4411529	3	US-09-103-840A-1
6	504	14.8	1704	1	US-08-103-840A-2
7	504	14.8	1704	1	US-08-528-199-2
8	488	14.3	2889	1	US-08-537-002A-4
9	488	14.3	2889	3	US-08-863-010-4
10	488	14.3	2889	3	US-09-024-429-4
11	488	14.3	3600	1	US-08-537-002A-5
12	488	14.3	3600	3	US-08-863-010-5

13	488	14.3	3600	3	US-09-024-429-5
14	404.5	11.9	1728	4	US-09-107-532A-2079
15	317	9.3	1683	4	US-09-134-001C-316
16	314.5	9.2	1635	4	US-09-107-532A-482
17	313.5	9.2	1782	1	US-08-374-155A-13
18	313.5	9.2	1782	2	US-08-785-396-13
19	302.5	8.9	4776	2	US-08-852-401-1
20	301.5	8.8	2074	1	US-08-039-777-2
21	301.5	8.8	2074	1	US-08-611-361A-2
22	287	8.4	4784	4	US-09-634-238-24
23	283.5	8.3	1794	1	US-08-374-155A-11
24	283.5	8.3	1794	2	US-08-785-396-11
25	280	8.2	1704	2	US-08-374-155A-15
26	280	8.2	1704	2	US-08-785-396-15
27	277.5	8.1	3895	4	US-09-961-527-201
28	259	7.6	4105	4	US-09-634-238-182
29	251.5	7.4	1608	3	US-08-137-077-1
30	251.5	7.4	1650	4	US-09-107-532A-3186
31	248	7.3	1668	4	US-09-107-532A-1068
32	245.5	7.2	6749	4	US-08-961-527-84
33	244	7.2	1890	1	US-08-374-155A-1
34	244	7.2	1890	2	US-08-785-396-1
35	243.5	7.1	1803	1	US-08-374-155A-9
36	243.5	7.1	1803	2	US-08-785-396-9
37	234.5	6.9	1305	1	US-08-374-155A-2
38	234.5	6.9	1305	2	US-08-785-396-2
39	233	6.8	2100	1	US-07-927-316A-1
40	227	6.7	1734	1	US-08-551-437-4
41	227	6.7	1734	3	US-03-004-225-4
42	227	6.7	1734	3	US-08-084-346-4
43	227	6.7	1734	3	US-03-104-704-4
44	212	6.2	1692	4	US-09-107-532A-2582
45	212	6.2	1692	4	US-09-107-532A-2583

#### ALIGNMENTS

RESULT 1  
US-08-737-752A-1  
; Sequence 1, Application US/08737752A  
; Patent No. 6265635

GENERAL INFORMATION:

APPLICANT: Kossmann, Jens  
APPLICANT: Welsh, Thomas

TITLE OF INVENTION: DNA SEQUENCES ENCODING ENZYMES CAPABLE

TITLE OF INVENTION: OF FACILITATING THE SYNTHESIS OF LINEAR  
TITLE OF INVENTION: ALPHA-1,4 GLUCANS IN PLANTS, FUNGI AND

TITLE OF INVENTION: MICROORGANISMS

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: James F. Haley, Jr., c/o Fish & Neave

STREET: 1251 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: United States of America

ZIP: 10020

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08737,752A

FILING DATE:

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

PRIOR APPLICATION NUMBER: DE P 44 17 879.4

FILING DATE: 18-MAY-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: DE P 44 47 388.5

FILING DATE: 22-DEC-1994

ATTORNEY/AGENT INFORMATION:

NAME: James F. Haley, Jr., Esq.  
REGISTRATION NUMBER: 27,794  
REFERENCES/DOCKET NUMBER: GFB-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 596-9000  
TELEFAX: (212) 596-9090  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2914 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Neisseria polysacchara  
IMMEDIATE SOURCE:  
LIBRARY: genomic library in pBluescriptII SK  
CLONE: pNB2  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 957..2867  
US-08-737-752A-1

## Alignment Scores:

Pred. No.: 0 Length: 2914  
Score: 3408.00 Matches: 636  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 3 Gaps: 0

US-09-843-007A-2 (1-636) x US-08-737-752A-1 (1-2914)

QY	1	MetLeuThrProThrGlnGlnValGlyLeuLeuLeuGlnTyrLeuLysThrArgIleLeu	20
DB	957	ATGTTGACCCCAACGAGCAAGTCGGTTGATTTTACGTACTCAAAACAGCATCTTG	1016
QY	21	AspIleTyrThrProGlnGlnArgAlaGlyIleGluLysSerGluAspTrpArgGlnPhe	40
DB	1017	GACATCTACACGCCGCAACAGCGCGCGCATCGAAATTCGAAAGACTGGCGGCGAGTTT	1076
QY	41	SerArgArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGly	60
DB	1077	TCGCGCGCATGATGATACGCAATTTCCCAACATGATGACGACTCGACAGCGTGTACGGC	1136
QY	61	AsnAsnGluAlaLeuLeuProMetLeuGluMetLeuLeuAlaGlnTyrGlnSerTyr	80
DB	1137	AACACGAGCCCTGCTGCTATGCTGGAATGCTGCTGCGCAGCATGGCAAGCTAT	1196
QY	81	SerGlnArgAsnSerSerLeuLysAspIleAspIleAlaArgGluAsnProAspTrp	100
DB	1197	TCCCAACGCAACTCATCTTTAAAGATATCGATATCGCGCGGAAACACCCCGATTGG	1256
QY	101	IleLeuSerAsnLysGlnValGlyValCysTyrValAspLeuPheAlaGlyAspLeu	120
DB	1257	ATTGTGTCACCAACCAAGTCGCGGCGTGTGTGCTGATTTGTTGCGCGCATTTG	1316
QY	121	LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrIleuHisLeu	140
DB	1317	AAGGCGTGAAGATAAATTCCTATTTTCAAGAGCTTGGTTGACTATATCTGCACCTG	1376
QY	141	MetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAlaValSerSerTyr	160
DB	1377	ATGCCGCTGTTAATATGCCCTGAAGGCAAAAGCAGCGCGCTATGCGGTACACAGCTAC	1436
QY	161	ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaLeu	180
DB	1437	CGCGATGCAATCGGCACCTGGGCACATAGTGGGACTTGGCGAAGTCAATTGTGCGCTG	1496
QY	181	HisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSerAsnGluHis	200

DB	1497	CACGAAGCCGGCATTTCCGGCGTGTGCGATTTTATCTTCAACACACACCTCCACGAACAC	1556
QY	201	GluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePhe	220
DB	1557	GAATGGGCGCAACGCTGCGCGCGCGCGCGCTTTTTCGACAATTTCTACTATAATTTC	1616
QY	221	ProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGln	240
DB	1617	CCCGACCGCGGATGCCGCAATACGACCGCACTGCGCGAATCTTCCCGGACGAG	1676
QY	241	HisProGlyGlyPheSerGlnLeuGluAspGlyArgTrpValTrpThrThrPheAsnSer	260
DB	1677	CACCGCGCGCTTCTCGCAACTGGAAGACGCGCTGGGTGTGAGACACCTTCAATTC	1736
QY	261	PheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheArgAlaMetAlaGlyGluMet	280
DB	1737	TTCCAATGGGACTTGAATTACAGCAACCCGCGGTATTTCCGCGCAATGGCGGCGAAATG	1796
QY	281	LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTrp	300
DB	1797	CTGTTCTTCCCAACTTGGCGGTGATGATCTTGGGTATGATGCGGTGCTTTATTGCG	1856
QY	301	LysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPhe	320
DB	1857	AAACAAATGGGACAGCTGCGAAACCTGCGCAGCGCGACGCCCTCATCGCGCGTTC	1916
QY	321	AsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHis	340
DB	1917	AATGCCGTATGCGTATTTGCCGCGCGCGCGGTCTTCTCAAAATCGAAGCATGTCGCAC	1976
QY	341	ProAspGlnValValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnProLeu	360
DB	1977	CCGACCAAGTCTGTCATATCATCGGCGAGGACGATGCGCAATCGGTTTACCAACCCCTG	2036
QY	361	GlnMetAlaLeuLeuTrpAsnThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAla	380
DB	2037	CAAAATGGCATTTGTTGTGAAACACCTTTGCCACGCGCGAAGTCAACCTGCTCCATCAGCG	2096
QY	381	LeuThrTyrArgHisAsnLeuProGluHisThrAlaTrpValAsnTyrValArgSerHis	400
DB	2097	CTGACTACCCGCAACACTGCGCGAGCATACCGCTGGGTCTCACTAGCTCCGACGCCAC	2156
QY	401	AspAspIleGlyTrpThrPheAlaAspGluAspAlaIleTyrLeuGlyIleSerGlyTyr	420
DB	2157	GACGACATCGGCTGCGCTGTCGCGATGAGACGCGCGCATATCTCGGGCATTAAGCGGCTAC	2216
QY	421	AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg	440
DB	2217	GACCAACCGCAATTTCTTCAACCGCTTCTTGTCAACCGTTTTCGACGCGAGCTTCGCTGT	2276
QY	441	GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla	460
DB	2277	GGCGTACCGTTCCATAACACCAAGCACAGCGACTGCGGTGTCAGTGTGACAGCCGCG	2336
QY	461	AlaLeuValGlyLeuAlaGlnAspAspProHisAlaValAspArgIleLysLeuLeuTyr	480
DB	2337	GCATTTGGTGGCGTGGCGCAAGACGATCCCAACCGCTTGACCCGCTCAAACTCTGTAC	2396
QY	481	SerIleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr	500
DB	2397	AGCATTTGCTTTGAGTACCGCGGCTGTCGCGTGTGATTTTACCTAGGCGACGAGTGGGTACG	2456
QY	501	LeuAsnAspAspAspTrpSerGlnAspSerAsnLysSerAspAspSerArgTrpAlaHis	520
DB	2457	CTCAATGACGACGACTGGTTCGCAACAGCAATTAAGACGACGACGCGGTGGCGGCGAC	2516
QY	521	ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGly	540
DB	2517	CCTCCGCGCTACACGACGCGCTGTACGCGCAACGCGACGATCCGTCGACCGCGCGGG	2576
QY	541	GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp	560
DB	2577	CAAACTCTATCAGGCTTGGCGCATATGATTTGCGCTCCGCAAAAGCAATCCGCGCTTCGAC	2636

QY 561 GlyArgGluValThrPheAsnThrAsnAsnLysHisIleLeuGlyTyrIleArgAsn 580  
DB 2637 GCGCGAGCGTGGTACATTCACCAACCAACAGCAACATCGCTACATCCGCAAC 2696  
QY 581 AsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThr 600  
DB 2697 AATGCGCTTTGGCATTCGGTAATTCAGCAATATCGCAAAACGGTTACCGCGCATACC 2756  
QY 601 LeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyLysThrValSerLeuAsn 620  
DB 2757 CTGCAAGCCATGCCCTTCAAGCGCGCACCGACCTCATCGGTGGGCAAACTGTGCGCTGAAT 2816  
QY 621 GlnAspLeuThrLeuGlnProTyrGlnValMetTyrPheLeuGluIleAla 636  
DB 2817 CAGGATTGACCGTTTCAGCCCTATCAGGTTCATGTGCTCGAATCGCC 2864  
RESULT 2  
US-09-252-991A-8258  
; Sequence 8258, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 8258  
; LENGTH: 3414  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-8258  
Alignment Scores:  
Pred. No.: 4,01e-55 Length: 3414  
Score: 543.50 Matches: 169  
Percent Similarity: 41.59% Conservative: 93  
Best local Similarity: 26.83% Mismatches: 227  
Query Match: 15.95% Indels: 141  
DB: 4 Gaps: 26  
US-09-843-007A-2 (1-636) x US-09-252-991A-8258 (1-3414)  
QY 85 SerSerLeuLysAspIleAspIleAlaArgGlu-----AsnAsnPro 98  
DB 91 TCGACGATAAGGAGAAACCCATGGCGCAGACCGGAGAACCGGTCTTTCTCAACGACCG 150  
QY 99 AspTyrIleLeuSerAsnLysGlnValGlyGlyValCysTyr-----ValAspLeu 115  
DB 151 CAATGGTACAGGAC-----CGCGTGATCTACCGGTCCACGCTGAATCC 195  
QY 116 Phe-----AlaGlyAspLeuLysGlyLeuLysAspIlePro 128  
DB 196 TTCTACGACGCCAACCAACGATGTATCGCGCATTCGCGCGCTGATCGAGAGCTCGAC 255  
QY 129 TyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeuMetProLeuPheLysCysProGlu 148  
DB 256 TACATCGCGACCTCGCGGTGACACCTCTCGGTGCTGCGCTCTCTAC-----CGGTCC 309  
QY 149 GlyLysSerAspGlyTyrAlaValSerSerTyrArgAspValAsnProAlaLeuGly 168  
DB 310 CCACGCGCGCACGCGCTACGACATCGCCCTAGTACCGTGGCGTGCACACGACGCTACGCG 369  
QY 169 ThrIleGlyAspLeuArgGluValIleAlaAlaLeuHisGluAlaGlyIleSerAlaVal 188  
DB 370 AGCTCGCGCAGCGCCCGCGCTTCATCGCCAGGCCCATCGACCGGTCTCGGGTGATT 429

QY 189 ValAspPheIlePheAsnHisThrSerAsnGluHisGluTrp-----AlaGlnArg 205  
DB 430 ACCGAGTGGTGATCAACCAACCACTCCGACGAGCATCCCTGGTTCATCCGCGCGCCGAC 489  
QY 206 CysAlaIaGlyAspProLeuPheAspAsnPheTyrTyrIlePheProAspArgMet 225  
DB 490 GCGAAGAGGATCGCGCGCGCGAC-----TACTACGTCTGTGTGCGAC-----AGC 537  
QY 226 ProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGlnHisProGlyGlyPhe 245  
DB 538 CACGAGAAATACCAAGGTAGC---CGGATCATCTTTCATCGAC----- 576  
QY 246 SerGlnLeuGluAspGlyArgTyrValTyrThr----- 256  
DB 577 -----ACCGAGCAGTCCCACTGGACCTGGGACCGCGTAGCCCAACAGTACTACTGGGAC 630  
QY 257 ThrPheAsnSerPheGlnTyrAspLeuAsnTyrSerAsnProTyrValPheArgAlaMet 276  
DB 631 CGTTCTTATTCACCGACCGGACCTGACTTCGACACCGCGAGGTCTCTCGCGAGGTG 690  
QY 277 AlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaVal 296  
DB 691 CTCGGGTGATCGCTACTGTGTCGACATGGCGCTCGACCGCTTCGCGCTCGACCGGATT 750  
QY 297 AlaPheIleTyrLysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeu 316  
DB 751 CCTTACTGTGACGACCGGACCGACCGAGCGAGAACCTTCGCGGAGACCCACCGAGGTG 810  
QY 317 IleArgAlaPheAsnAlaValMetArgIleAlaProAlaValPhePheLysSerGlu 336  
DB 811 CTCAGGCGCATCGCGCGCGAGCTGTCGCGGACCTATCCGCGACCGCATCTGCTGCGCGAG 870  
QY 337 AlaIleValHisProAspGlnValValGlnTyrIleGly-----Gln 350  
DB 871 GCCAACCGAGTGGCGGAGAACACCGCGCGTACTTCGCGCGGAGGATGGCGGCGAGGCG 930  
QY 351 AspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTyrAsnThrLeuAla 370  
DB 931 GACGATGCGACATGGCTTCCACTCCCGCTGATGCGCGCGCATATACATGCCATCGCC 990  
QY 371 ThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHis 390  
DB 991 CAGGAGGATTCGCTATCCGATCACCGACATCTTCGCGCAGACCCCGGACATCCCGCCAT 1050  
QY 391 ThrAlaTyrValAsnTyrValArgSerHisAspIleGlyTyrThrPheAlaAspGlu 410  
DB 1051 TGCCATGGCGGATCTTCTCGCGCAACCGACCGAGCTGACCTGGAGATGTCACCGAC 1110  
QY 411 AspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsnArgPhePhe 430  
DB 1111 -----GACGAGCGCGACTATCTC----- 1128  
QY 431 ValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsnProSerThr 450  
DB 1129 -----TGAAACCATATGCC 1143  
QY 451 GlyAspCysArgValSer-----GlyThrAlaAlaAlaLeuValGlyLeuAlaGln 467  
DB 1144 GCCGACCGCGCGCGCTCAACCTGGGCTCCCGCGCGCTCGCGCGCTGTGGAG 1203  
QY 468 AspAspProHisAlaValAspArgIleLysLeuTyrSerIleAlaLeuSerThrGly 487  
DB 1204 CGTGACCGCGCG-----CGCATCGAGTGTGTCACACGCTCTGTGTTCGATCGCG 1254  
QY 488 GlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr-----LeuAsn 502  
DB 1255 GGCACGCGCGACCTGTACTACGCGCGAGATCGGATCGGCGGACCAACATCTACCTCGGC 1314  
QY 503 AspAspAsp-----TyrSerGlnAspSerAsn----- 511  
DB 1315 GACCGCGACGCGGTGCGGACCCCGATGTCAGTGTGTGTCGCGGACCGCGCGGTCTTCC 1374  
QY 512 LysSerAspAspSerArgTyrAlaHisArgProArgTyrAsnGluAlaLeuTyr----- 529

```

Db      1375  CGGCGGACCGCGGAGCTGGTG---CTGCGCGCGATCTCGACCGCTGTACGGCTAC 1431
Qy      530  -----AlaGlnArgAsnAspProSerThrAlaAlaGlyGlnIleTyr 543
Db      1432  CAGACGATCAACGTGAGGCGGAGCGCGACCGGATTGG-----CTGCTC 1479
Qy      544  GlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAspGlyArg 563
Db      1480  AACTGGATGCGCGCTGCTGCGCGTGGCGGAGCAGCAGAGGCGCTTCGCGCGCGGCGAGC 1539
Qy      564  LeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIle----- 578
Db      1540  CTGAAGATGTCGCGCGCGGAGCAACGAGGATCTCCCTACCTGCGCGGATAGCGGAG 1599
Qy      579  -----ArgAsnAsnAlaLeuLeuAlaPheGlyAsnPheSerGlyPyrProGlnThrVal 596
Db      1600  GCGAAGCGGAGGACAGATCTCTGCGTGCGCAACCTGTCGCGCGCGCGCGCGGCGTG 1659
Qy      597  ThrAlaHisThrLeuGlnAlaMetProPheLysAlaHis-----Asp 610
Db      1660  -----GAACTGGACCTCGCCAGCCATCGCGGCAAGGTGCGCGGTGGAG 1701
Qy      611  LeuIleGlyGlyLysThrVal-----SerLeuAsnGlnAspLeuThrLeuGln 626
Db      1702  ATGATCGCGCGCATGTGTTCCCGCGCATCGCGGAGCTGACCTACCTGCTGACCTGCGG 1761
Qy      627  ProTyrGlnValMetTyrLeuGluIleAla 636
Db      1762  CCTACGCGCTTCTACTGCTTCTACCTGCGC 1791

RESULT 3
US-09-252-991A-8065/c
; Sequence 8065, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8065
; LENGTH: 4125
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8065

Alignment Scores:
Pred. No.: 5,536-55 Length: 4125
Score: 543.50 Matches: 169
Percent Similarity: 41.59% Conservative: 93
Best Local Similarity: 26.83% Mismatches: 227
Query Match: 15.95% Indels: 141
DB: 4 Gaps: 26

US-09-843-007A-2 (1-636) x US-09-252-991A-8065 (1-4125)
Qy      85  SerSerLeuLysAspIleAspIleAlaArgGlu-----AsnAsnPro 98
Db      3801  TCGAGCGCATAAGGAGAAACCCATGCCAGCGCGAGAAACCGGCTTTCTTCAACGACCGG 3742
Qy      99  AspTyrIleLeuSerAsnLysGlnValGlyGlyValCysTyr-----ValAspLeu 115
Db      3741  CAATGGTACACAGGAC-----CGGTTGATCTACAGGTCCAGCTGAATCC 3697
Qy      116  Phe-----AlaGlyAspLeuLysGlyLeuLysAspLysIlePro 128

```

```

Db      3696  TTCTACGAGCGCAACACGATGGTATCGCGGACTTCGCGGCGGTGATCGAGAGCTCGAC 3637
Qy      129  TyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeuMetProLeuPheLysCysProGlu 148
Db      3636  TACATCGCGCGACCTCGGCGTGAACACTCTCTGCGTGTCTGCGGTCTAC-----CGGTGG 3583
Qy      149  GlyLysSerAspGlyGlyTyrAlaValSerSerTyrArgAspValAsnProAlaLeuGly 168
Db      3582  CCACGCCGCGAGCGGTCTACGATCGCGCGGAGTACCGGTGCGGTGCGGTGCGGTGCGGT 3523
Qy      169  ThrIleGlyAspLeuArgGluValIleAlaAlaLeuHisGluAlaGlyIleSerAlaVal 188
Db      3522  AGCTTCGCGCGCGCGCGGTTCATCGCGGAGGCCATCGACCGGTCTGCGGTGAT 3463
Qy      189  ValAspPheIlePheAsnHisThrSerAsnGluHisGluTyr-----AlaGlnArg 205
Db      3462  ACCGAGCTGGTATCAACACACCTCGACGAGTCCCTGGTTCATCCGCGCGCGCGAC 3403
Qy      206  CysAlaAlaGlyAspProLeuPheAsnPheTyrTyrIlePheProAspArgMet 225
Db      3402  GCGAAGAGGATCGCGCGCGCGCGAC-----TACTACGTCTGTCGCGAC-----AGC 3355
Qy      226  ProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGlnHisProGlyPhe 245
Db      3354  GACGAGAAATACCGAGGTACG---CGGATCATCTTCATCGAC----- 3316
Qy      246  SerGlnLeuGluAspGlyArgTyrValTyrThr----- 256
Db      3315  -----ACGAGCATGTCACACTGGACCTGGACCGCGGTAGCCCAACAGTACTACTGGCAC 3262
Qy      257  ThrPheAsnSerPheGlnTyrAspLeuAsnTyrSerAsnProTyrPheArgAlaMet 276
Db      3261  CGCTTCTATTCCACCGCGGACCTGAACTTCGACACCGCGAGTCTCTGCGCGAGGTG 3202
Qy      277  AlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaVal 296
Db      3201  CTGCGGCTGATGCGCTACTGCGTGGACATGGCGGTGCGGTGCGGTGCGGTGCGGTGAT 3142
Qy      297  AlaPheIleTyrLysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeu 316
Db      3141  CCCTACTGATCGAACCGGACGCGACCGACGAGAGAACCTGCGCGAGACCCACAGGTG 3082
Qy      317  IleArgAlaPheAsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGlu 336
Db      3081  CTCACGCGCATCGCGCGCGAGCTGGAGCGCGCTATCCGACCGCGATGCTGCTGCGCGAG 3022
Qy      337  AlaIleValHisProAspGlnValValGlnTyrIleGly-----Gln 350
Db      3021  GCCAACCGTGGCGGAGACACCGCGCGGTACTTCGCGCGGAGGATGGCGCGAGGGC 2962
Qy      351  AspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuThrPheAsnThrLeuAla 370
Db      2961  GACGAAATGCCACATGGCGCTTCCACTTCCGCTGATCCGCGCATGTACATGGCCATCGCC 2902
Qy      371  ThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHis 390
Db      2901  CAGGAGATGCTATCCGATCCCGACATCTCGCGCGAGACCCCGCGACATCCCGGCAAT 2842
Qy      391  ThrAlaTyrValAsnTyrValArgSerHisAspAspIleGlyTyrThrPheAlaAspGlu 410
Db      2841  TGCCAATGGCGCATCTTCTGCGCAACACGAGAGCTGACCTGGAGATGGTCAACCGAC 2782
Qy      411  AspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsnArgPhePhe 430
Db      2781  -----GACGAGCGGACATCTCTC----- 2764
Qy      431  ValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsnProSerThr 450
Db      2763  -----TGGACCACTATGCC----- 2749
Qy      451  GlyAspCysArgValSer-----GlyThrAlaAlaAlaLeuValGlyLeuAlaGln 467
Db      2748  GCGAGCGCGCGCGCGCGCTCAACCTGGGCGATCCGCGCGCGCGCGCGCTGCGGTGGAG 2689

```

QY	468	AspAspProHisIaValIaAspArgIleLysLeuLeuTyrSerIleAlaLeuSerThrGly	487
Db	2688	CGTGCACGGCGG-----CGCATCGAGCTGCTGCACAGCCTGCTGTTGTGTCAGTCCG	2638
QY	488	GlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr-----LeuAsn	502
Db	2637	GGCAGCCGACCTGTACTACGGCGAGAGATCGGATCGGCGACACATCTACCTCGGC	2578
QY	503	AspAspAsp-----TrpSerGlnAspSerAsn-----	511
Db	2577	GACCGCGAGCGGCTGGCGACCCGATGCAGTGGTCTGGTGCACCGCAACGCGGCTTCTCC	2518
QY	512	LysSerAspAspSerArgTrpAlaHisArgProArgTyrAsnGluAlaLeuTyr-----	529
Db	2517	CGCGCCGACCCGCGAGCTGGTG-----CTGCCCGCATCTCTGCACCGCTGTACGGCTAC	2461
QY	530	-----AlaGlnArgAsnAspProSerThrAlaAlaGlyGlnIleTyr	543
Db	2460	CACAGCATCAACGTCAGGCGGACCGCGCGCATTCG-----CTGCTC	2413
QY	544	GlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAspGlyIleArg	563
Db	2412	AACGTGATCGCGCTGCTGCGCGTGGCGAGCCAGCAGAGAGCCCTCGGCGCGCGCGG	2353
QY	564	LeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIle-----	578
Db	2352	CTGAAGATGCTCGCGCGAGCAACCGAGGATCTCGCTACTGCGCGAATACGCGGAG	2293
QY	579	-----ArgAsnAsnAlaLeuAlaPheGlyAsnPheSerGluTyrProGlnThrVal	596
Db	2292	GGCGAACGGCAGACAGCATCTCTCGTGGCGCAACCTGTCTGCGCGCGCGCGCGG	2233
QY	597	ThrAlaHisThrLeuGlnAlaMetProPheLysAlaHis-----Asp	610
Db	2232	-----GNACTGGACCTCGCGAGCATGCCGCGCAAGTGCCTGGTGGAG	2191
QY	611	LeuIleGlyIleLysThrVal-----SerLeuAsnGlnAspLeuThrLeuGln	626
Db	2190	ATGATCGGCGCATGCTGCTCCCGCGATCGCGAGCTGACCTACCTGTGACCTCGCG	2131
QY	627	ProTyrGlnValMetTrpLeuGluIleAla	636
Db	2130	CCCTACGGCTTCTACTGTTCTACCTCGGCC	2101
RESULT 4			
US-09-103-840A-2			
; Sequence 2, Application US/09103840A			
; Patent No. 6294328			
; GENERAL INFORMATION:			
; APPLICANT: FLEISCHMAN, Robert D.			
; APPLICANT: WHITE, Owen R.			
; APPLICANT: FRASER, Claire M.			
; APPLICANT: VENTER, John C.			
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM			
; FILE OF INVENTION: TUBERCULOSIS			
; FILE REFERENCE: 24366-20007.00			
; CURRENT APPLICATION NUMBER: US/09/103,840A			
; CURRENT FILING DATE: 1998-06-24			
; NUMBER OF SEQ ID NOS: 2			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 2			
; LENGTH: 4403765			
; TYPE: DNA			
; ORGANISM: Mycobacterium tuberculosis			
; FEATURE:			
; OTHER INFORMATION: CDC 1551			
; OTHER INFORMATION: "n" bases at various positions throughout the sequence			
; OTHER INFORMATION: represent a, t, c or g			
US-09-103-840A-2			
Alignment Scores:			
Pred. No.:		1.05e-45	Length: 4403765



```

RESULT 5
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

Alignment Scores:  
 Pred. No.: 1,05e-45 Length: 4411529  
 Score: 509.50 Matches: 160  
 Percent Similarity: 41.90% Conservative: 83  
 Best Local Similarity: 27.59% Mismatches: 216  
 Query Match: 14.95% Indels: 121  
 DB: 3 Gaps: 24

US-09-843-007A-2 (1-636) x US-09-103-840A-1 (1-4411529)

QY	117	AlaGlyAspLeuLysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThr	136
DB	152514	TCGGCGGATCTCGTGGAGTCTATCGATCGCTCGACTACTCGAGTGGCTTGGCATCGAC	152573
QY	137	TyrLeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAla	156
DB	152574	TGCATCTGGTTCGGCCGCTTACGACTCGCG-----CTGGCGGACGGCGGTTCACG	152627
QY	157	ValSerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluVal	176
DB	152628	ATTGCGCATCTTACCAAGGTGCTGCCGGAATTCCGACACCGTACACGATTTTCGTCCGCCCT	152687
QY	177	IleAlaAlaLeuHisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThr	196
DB	152688	GTGACGCGCGTACCAGGAGGTATCCGCATCATCCGACTTGGTGTGATGAATCAGACC	152747
QY	197	SerAsnGluHisGluTrpAlaGln-----ArgCysAlaAlaGlyAspProLeuPheAsp	214
DB	152748	TCGGAGTCGCACCCCTGGTTTCAGGAGTCCGCGCGGACCCACGACGACCGTACGGTGAC	152807
QY	215	AsnPheTyrTyrIlePheProAsp-----ArgArgMetProAspGlnTyrAspArgThr	232
DB	152808	-----TATPAGCTGGAGCGACACAGCGAGCGCTTACCCGAC-----	152846
QY	233	LeuArgGluIlePheProAspGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArg	252
DB	152847	GCCCGGATCATCTTCGTCGAC-----valTrpThrPheAsnSerPheGlnTrp	152882
QY	253	Trp-----valTrpThrPheAsnSerPheGlnTrp	263
DB	152883	TGTCATATTCGATCCTGTCGCGGACAGTTCCTACTGGACCGATTCCTCTCCACCAACCG	152942
QY	264	AspLeuAsnTyrSerAsnProTyrValPheArgAlaMetAlaGlyGluMetLeuPheLeu	283
DB	152943	GATCTGAATACGACAACCCCGCGTGAAGAGCGATGATCGACGTCATCCGCTTTTGG	153002
QY	284	AlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTyrLysGlnMet	303
DB	153003	CTCGGCTGGGCATCGACGGGTTTCGGTGTGGACCGGTGCCCTTCTTTGAACCGTGA	153062
QY	304	GlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPheAsnAlaVal	323
DB	153063	GGACCAACTGCGAGAACCTTCGCGGAACACACACGCTTTCTCAAGCGAGTCCGCAAGTG	153122
QY	324	MetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHisProAspGln	343
DB	153123	GTGGACGACGAATTCGCCGCGCGGTGCTGTAGTCCGAACCAATAGTGGCCGGCGAT	153182
QY	344	ValValGlnTyrIleGlyGln-----AspGluCysGlnIleGlyTyrAsn	358
DB	153183	GTGTCGATATTCGTTGATCCCAACACCGGTGGCGAGTGCACATGGCCCTTTTTCAC	153242
QY	359	---ProLeuGlnMetAlaLeuLeuTyrAsnThrLeuAlaThrArgGluValAsnLeu---	376
DB	153243	TTCCCGCTGATCGCGGCACTTTC-----ATGCGCGTGGCGCGGAGTCCCGTTTT	153293
QY	377	---LeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHisThrAlaTyrValAsn	395
DB	153294	CGGATCTCGGAGATCATCGCCAGACCCCAACCATCTGCATGGCGCAATGGGGATA	153353
QY	396	TyrValArgSerHisAspAspIleGlyTrpThrPhe---AlaAspGluAspAlaTyr	414

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/528,199
FILING DATE:
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/485,126
FILING DATE: 14-SEP-1995
APPLICATION NUMBER: US 08/485,126
FILING DATE: 07-JUN-1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: JP 156399/1994
FILING DATE: 16-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: KUBOTA-6A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1704 base pairs
TYPE: nucleic acid
STRADEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-528-199-2

Alignment Scores:
Pred. No.: 7,71e-51 Length: 1704
Score: 504.00 Matches: 158
Percent Similarity: 41.89% Conservative: 95
Best Local Similarity: 26.16% Mismatches: 232
Query Match: 14.79% Indels: 120
DB: Gaps: 22

US-09-843-007A-2 (1-636) x US-08-528-199-2 (1-1704)

QY 98 ProAsnTyrLeu-----LeuSerAsnLys 105
Db 22 CCGAGTGGTTCGCGACGGCGGTCTTCTACGAGTCTGTGTGGTCTCTCCGGGACCCC 81
QY 106 GlnValGlyValCysTyrValAspLeuPheAlaGlyAspLeuLysGlyLeuLysAsp 125
Db 82 AAGCGCGGGGCG-----ACGGGTGACTCCGCGGCGCTCGCGGAG 120
QY 126 LysIleProTyrPheGlnLeuGlyLeuThrTyrLeuHisLeuMetProLeuPheLys 145
Db 121 AAGCTCGACTACCTGAGTGGCTCGGCGTCCGACTGCGTGTGGGTGCGCGCGTCTTTCAGC 180
QY 146 CysProGluGlyLysSerAspGlyGlyTyrAlaValSerTyrArgAspValAsnPro 165
Db 181 TCGGCG-----CTGGCGGACGGGGGTACAGCTCGCCACTACACGGGATCTCCCG 234
QY 166 AlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaLeuHisGluAlaGlyIle 185
Db 235 GAGATCGGCACGTCGAGGACTTCCAGCGCTCTCTCGACGGCGGCACGAGCGGGGATC 294
QY 186 SerAlaValAspPheIlePheAsnHisThrSerAsnGluHisGluTyr-----Ala 203
Db 295 CGGGTGATCATCGATTGTCATGAACACACAGAGTACCGCGCACCGGCTTCAGGCCC 354
QY 204 GlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePheProAspArg 223
Db 355 TCCCGCAGCGATCCGACGGCGCGTACGGCGAC-----TTCTACGCTCTGTCTCGAC 405
QY 224 ArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAspGlnHisProGly 243
Db 1000

```

```

1  RESULT 6
2  US-08-528-199-2
3  ; Sequence 2, Application US/08528199
4  ; Patent No. 5763228
5  ;
6  ; GENERAL INFORMATION:
7  ;
8  ; APPLICANT: KUBOTA, Michio
9  ; APPLICANT: TSUSAKI, Keiji
10 ; APPLICANT: SUGIMOTO, Toshiyuki
11 ; TITLE OF INVENTION: RECOMBINANT ENZYME FOR CONVERTING
12 ; TITLE OF INVENTION: MALTOSE INTO TREHALOSE
13 ; NUMBER OF SEQUENCES: 8
14 ;
15 ; CORRESPONDENCE ADDRESS:
16 ; ADDRESSEE: BROWDY AND NEIMARK
17 ; STREET: 419 Seventh Street, N.W., Suite 300
18 ; CITY: Washington
19 ; STATE: D.C.
20 ; COUNTRY: USA
21 ; ZIP: 20004
22 ;
23 ; COMPUTER READABLE FORM:
24 ;

```

Db 406 -----ACCGACGAGCTCTACACGACGCGCGGGTGTACTTCTCGTACACACGACGCGGTGC 459  
Qy 244 GlypheSer---GlnLeuGluAspGlyArgTrpValTrpThrPheAsnSerPheGln 262  
Db 460 AACTGGAGCTGGGACACGACCGCGCGGCTACTACTGCGACCGCTTCTTCACACACGAG 519  
Qy 263 TrpAspLeuAsnTySerAsnProTrpValPheArgAlaMetAlaGlyGluMetLeuPhe 282  
Db 520 CCGGACCTGAATTCGACACCGGAGGTCAGGACGCTCTCGAGCGGATGCGGTTC 579  
Qy 283 LeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTrpGln 302  
Db 580 TGGCTCGACATGGGCTCGAGCGCTTCGGCTCGACGGGTCCCTACTCTACGAGGT 639  
Qy 303 MetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPheAsnAla 322  
Db 640 CCGGACACAGCGGAGAGACCTCCCGGAGCGGACGACGAGATCTCAAGGGGTGGCGGC 699  
Qy 323 ValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHisProAsp 342  
Db 700 TTCGTGACGACACACTACCCGACCGGGTGTCTGTACGAGCGGAAACGAGTGGCGAC 759  
Qy 343 GlnValValGlnTyIleGlnAspGluCysGlnIleGly----- 356  
Db 760 GACGTGGTGGATCTTCGGGCGGAGGAGCGTGGAGCGGACCGGTCTCGGGCCCGAG 819  
Qy 357 -----TyrAsnProLeuGln-MetAlaLeuLeuTrpAsnThrLeuAla 371  
Db 820 AGTCACATGGCTTCACCTTCGGGTGTCGCGCATCTTCATGCGGTGGCGCGGAG 879  
Qy 371 rArgGluValAsnLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHis 391  
Db 880 TGGC---GCTCCGATCTCGGAGATCATGAGCAGACCGCGC-ATCCCGGGGGCTG 935  
Qy 391 rAlaTrpValAsnTyValArgSerHisAspIleGlyTrpThrPhe---AlaAsp 410  
Db 936 CCAGTGGGCACTCTCTCGGACACACGACGAGTGCACCTCGAGATGGTCACGACGA 995  
Qy 410 uAspAlaAlaTyLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsnArgPhe 430  
Db 996 GCACCGCACTACATGG- 1014  
Qy 430 eValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyAsnProSer 450  
Db 1015 -----GGGAGTACGCCAGGAGACCCCGCATGAGGCCAAC- 1050  
Qy 450 rGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAlaGlnAspPr 470  
Db 1051 -----ATCGGATCGGCGCGGCTCGCGCTCTCGACACGAC- 1092  
Qy 470 chiAlaValAspArgIleGlyLeuLeuTySerIleAlaLeuSerThrGlyGlyLeuPr 490  
Db 1093 -----ACGACACGATCGAGTGTTCACCGGCTGTGCTGCTCGCGGTCCCGTCC 1145  
Qy 490 oLeuIleTyLeuGlyAspGluValGlyThrLeuAsnAspAspTrpSerGlnAspSe 510  
Db 1146 GTCTCTGTACTACGGGACAGATCGC---ATGGCGACACATCTGCTCGGTGACCG 1202  
Qy 510 rAsn-----LysSerAspAspSerArgTrpAla----- 519  
Db 1203 CGACGGCGTGCGTACCGCATCGAGCGACCCCGACCGCATCGGCTTCTCGCGCGAC 1262  
Qy 520 -----HisArgProArgTyArgGlnAlaLeuTy----- 529  
Db 1263 CAGCGCGGACAGCTCATCTGCGGACGATCCAGGACCGGCTACGGCTACGAGCGT 1322  
Qy 530 -----AlaGlnArgAsnAspProSerThrAlaAlaGlyGlnIleTyGlnGly 546  
Db 1323 CAAGTTCGAGGCGCAGCTGGAGAACCCCTCTCG-----CTGCTGCATCTGGAC 1370  
Qy 546 uArgHisMetIleAlaValArgGlnSerAsnProArgPheAspGlyGlyArgLeuVal 566  
Db 1371 CCGCGCGATGATCCATCCGCGCGGACCGGACGCGCTTCGGGCTGGCAGCACTTCAGGA 1430

Qy 566 rPheAsnThrAsnAsnLysHisIleIleGlyTyIleArg----- 579  
Db 1431 CTCCGGCGGTTCGAACCCCGCGGTGCTCTAGTGGCGAGCTGCCGGCGACGGGG 1490  
Qy 580 -AsnAsnAlaLeuLeuAlaPheGlyAsnPheSerGlyTyProGlnThrValThrAlaHi 599  
Db 1491 CGACGAGTGTCTCTCTCGGTCAACACCTGTCCCGTTCGCGACGCGGTC----- 1542  
Qy 599 sThrLeuGlnAlaMetProPheLysAlaHis-----AspLeuIleGlyGly----- 614  
Db 1543 -GAGCTCGACCTCCGGAAGTACGAGGCGCGGTACCGGTGAGCTGATCGCGCGGTGCC 1601  
Qy 615 -----LysThrValSerLeuAsnGlnAspLeuThrLeuGlnProTyArgGlnValMetTr 632  
Db 1602 GTTCCCGCGGTCCGGGAGCTCCGATATCTCTGACCTCAGCGGCGACGCGCTTCTACTG 1661  
Qy 632 pLeuGluIle 635  
Db 1662 GTTCCGGCTC 1671

RESULT 7  
US-08-528-199-5  
; Sequence 5, Application US/08528199  
; Patent No. 5763228  
; GENERAL INFORMATION:  
; APPLICANT: KUBOTA, Michio  
; APPLICANT: TSUSAKI, Keiji  
; APPLICANT: SUGIMOTO, Toshiyuki  
; TITLE OF INVENTION: RECOMBINANT ENZYME FOR CONVERTING  
; TITLE OF INVENTION: MALTOSE INTO TREHALOSE  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/08/528,199  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,126  
; FILING DATE: 14-SEP-1995  
; APPLICATION NUMBER: US 08/485,126  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 156399/1994  
; FILING DATE: 16-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BROWDY, Roger L.  
; REGISTRATION NUMBER: 25,618  
; REFERENCE/DOCKET NUMBER: KUBOTA-6A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1704 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS

LOCATION: 1..1704  
US-08-528-199-5

Alignment Scores:  
Pred. No.: 7,71e-51 Length: 1704  
Score: 504.00 Matches: 158  
Percent Similarity: 41.89% Conservative: 95  
Best Local Similarity: 26.16% Mismatches: 232  
Query Match: 14.79% Indels: 120  
DB: 1 Gaps: 22

US-09-843-007a-2 (1-636) x US-08-528-199-5 (1-1704)

Qy	98	ProAspTTPile-----	-----LeuSerAsnlys	105
Db	22	CCCGAGTGGTCCGACCGCGGTCTTCTACGAGGTCTGGTGGTCCCTTCGGGACCCC	81	
Qy	106	GlnValGlyGlyValCyeTyTzValAspLeuPheAlaGlyAspLeuLysGlyLeuLysA	125	
Db	82	AACGCGCGCGGCG-----	-ACGGGTGACTTCCGGCGGCTCGCGGAG	120
Qy	126	LysIleProTyPheGlnGluLeuGlyLeuThrTyTzLeuHisLeuMetProLeuPheLys	145	
Db	121	AAGCTCGACTACCTGCACTGCTCGCGTCTGACTGCTGCTGGTGGTCCGCTTCTTACG	180	
Qy	146	CysProGluGlyLysSerAspGlyTyTzAlaValSerSerTyTzArgAspValAsnPro	165	
Db	181	TCGCGG-----CTGCGGACCGGGGTACGAGGTGCGGACTACACGGGATCTCCCG	234	
Qy	166	AlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaLeuHisGluAlaGlyIle	185	
Db	235	GAGATGCGGACCGGTGAGGACTTCCACGCTTCTCGACGCGGCGGACGAGCGGGATC	294	
Qy	186	SerAlaValAspPheIlePheAsnHisThrSerAsnGluHisGluTrp-----	Ala	203
Db	295	CGGTGATCATCGACTTCGTATGATACACACGAGTGACGCGCACCCGCTGTTCCAGGCC	354	
Qy	204	GlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyTzIlePheProAspArg	223	
Db	355	TCCGCGACGATCCGACGCGCGGTACGCGGAC-----TTCTAGTGTGTCGAC	405	
Qy	224	ArgMetProAspGlnTyTzAspArgThrLeuArgGluIlePheProAspGlnHisProGly	243	
Db	406	-----ACCGACGAGCTTACGAGACCGCGCGGTGATCTTCTGACACCGAGCGGTG	459	
Qy	244	GlyPheSer-----GlnLeuGluAspGlyArgTrpValTrpThrPheAsnSerPheGln	262	
Db	460	AACGAGAGTGGGACCGACACCGCGCGGTGATCTTCTGACACCGAGCGGTG	519	
Qy	263	TrpAspLeuAsnTyTzSerAsnProTrpValPheArgAlaMetAlaGlyLeuMetLeuPhe	282	
Db	520	CCGACCTGAACTTCGACAAACCGAGGTCCAGGACGACCATGCTGGAGGCGATGGCGTTC	579	
Qy	283	LeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTrpLysGln	302	
Db	580	TGGCTCGACATGGGCTCGACGGTCTCGGCTCGACGCGGTGCTTCTACGAGCGT	639	
Qy	303	MetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPheAsnAla	322	
Db	640	CCCGGACCAACGCGGAGAACTCCCGAGACGACGAGATGCTCAAGCGGTGCGCGC	699	
Qy	323	ValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHisProAsp	342	
Db	700	TTCGTGACGACAACTACCCCGCGGTGCTGCTGACGAGGCGACCACTGCGCGGAC	759	
Qy	343	GlnValValGlnTyTzIleGlyGlnAspGluCysGlnIleGly-----	356	
Db	760	GACGTGGTGGATCTTCCGCGCGGAGGAGCGGTGAGGACGCGGTGCGCGCGGAG	819	
Qy	357	-----TyTzAsnProLeuGln-MetAlaLeuLeuTrpAsnThrLeuAlaTh	371	
Db	820	AGTCACATGGCTTCCACTTCCCGGTGATGCGCGGATCTTCTGAGGCGGCGGCTTCTACTG	879	

Qy	371	rArgGluValAsnLeuLeuHisGlnAlaLeuThrTyTzArgHisAsnLeuProGluHisTh	391	
Db	880	TCGC---GCTTCCGATCTCGAGATCGAGCAGACGCGCG-ATCCGAGAGGCTG	935	
Qy	391	rAlaTrpValAsnTyTzValArgSerHisAspAspIleGlyTrpThrPhe---AlaAspG	410	
Db	936	CCAGTGGGATCTTCTCGCAACACGACGAGTACCTCGAGATGGTCCACGACGA	995	
Qy	410	uAspAlaAlaTyTzLeuGlyIleSerGlyTyTzArgHisArgGlnPheLeuAsnArgPheP	430	
Db	996	GGACCGCGACTACATGG------	1014	
Qy	430	eValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyTzAsnProSerTh	450	
Db	1015	-----GGGAGTACCAAGGACCCCGCATGAGGCCAAC-----	1050	
Qy	450	rgLysPysArgValSerGlyThrAlaAlaLeuValGlyLeuAlaGlnAspAspPr	470	
Db	1051	-----ATCGCATCCGCGCGGCTCGCGCTGCTGCTGACACGAC--	1092	
Qy	470	oHisAlaValAspArgIleLysLeuLeuTyTzSerIleAlaLeuSerThrGlyGlyLeuPr	490	
Db	1093	-----ACGACCGATCGAGCTTTCACCGGCTCTGCTGCTGCGGCTCCCC	1145	
Qy	490	oLeuIleTyTzLeuGlyAspGluValGlyThrLeuAsnAspAspTrpSerGlnAspSe	510	
Db	1146	CGTCTGTACTACGCGACGAGATCGGC---ATGCGCGACACATCTGGCTGCTGACCG	1202	
Qy	510	rAsn-----LysSerAspAspSerArgTrpAla-----	519	
Db	1203	CGAGCGGTGCTGACGCGGATGACGCGGACCCCGACCGCACTCGGCTTCTCGCGCGC	1262	
Qy	520	-----HisArgProArgTyTzAsnGluAlaLeuTyTz-----	529	
Db	1263	CACGCGCGCAAGTCGACCTCGCGAGATCCAGGACCGGCTCTACGGCTACGAGCGT	1322	
Qy	530	-----AlaGlnArgAsnAspProSerThrAlaAlaGlyGlnIleTyTzGlnGlyLe	546	
Db	1323	CAACGTGACGCGCGAGCTGGAGAACCCCTCTCG------CTGTCGACCTGGAC	1370	
Qy	546	uArgHisMetIleAlaValArgGlnSerAsnProArgPheAspGlyGlyArgLeuValTh	566	
Db	1371	CGCGCGATGATCCACATCCGCGCGGACGCGGCTTGGGTCGGGACCTTCGAGGA	1430	
Qy	566	rPheAsnThrAsnAsnLysIleIleGlyTyTzIleArg-----	579	
Db	1431	CCTCGCGCGCTCGAACCCGCGGTGCTGCTCTACTGCTGCGGAGTCCGCGGCGAG	1490	
Qy	580	-AsnAsnAlaLeuAlaPheGlyAsnPheSerGluTyTzProGlnThrValThrAlaHi	599	
Db	1491	CGAGCGGTGATCTCTGGGTCAACACCTGTCCTGCTTCCGCGGCGGCTC-----	1542	
Qy	599	sThrLeuGlnAlaMetProPheLysAlaHis-----AspLeuIleGlyGly-----	614	
Db	1543	-GAGCTCGACCTCGGAAAGTACGAGGCGCGGTACCGGTGAGCTGATCGCGCGGTGCC	1601	
Qy	615	-----LysThrValSerLeuAsnGlnAspLeuThrLeuGlnProTyTzGlnValMetTr	632	
Db	1602	GTTCCCGCGGTGCGGAGCTCCGCTATCTCTGAGCTCAGCGGCGGCGGCTTCTACTG	1661	
Qy	632	pLeuGluIle	635	
Db	1662	GTTCCGCGCTC	1671	

## RESULT 8

US-08-537-002A-4  
Sequence 4, Application US/08537002A  
Patent No. 5773282  
GENERAL INFORMATION:  
APPLICANT: TSUSAKI, Keiji  
APPLICANT: KUBOTA, Michio  
APPLICANT: SUGIMOTO, Toshiyuki  
TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR



Qy 561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArgAsn 580  
Db 1369 CGGGAGCCCTACCTCTCCCGTGGAGAACCGCGCGCTCTCGCTACCTAGGGAG 1428  
Qy 581 Asn-----AlaLeuAlaPheGlyAsnPheSerGlyTyrProGlnThrValThr 597  
Db 1429 CACGAGGGGAGCGGCTCTGCTGGTGGCCAACTCTCCCGCTACACCGAGCCCTTT--- 1485  
Qy 598 AlaHisThrLeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyThrVal 617  
Db 1485 ----- 1485  
Qy 618 SerLeuAsnGlnAspLeuThrLeuGlnProTyrGln 629  
Db 1486 -----GACCTCCCTTGGAGCGCTTACCAA 1509

# RESULT 9

US-08-863-010-4  
Sequence 4, Application US/08863010  
Patent No. 6087146  
GENERAL INFORMATION:  
APPLICANT: TSUSAKI, Keiji  
APPLICANT: KUBOTA, Michio  
APPLICANT: SUGIMOTO, Toshiyuki  
TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR  
TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESS: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/863,010  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/537,002  
FILING DATE: 29-SEP-1995  
APPLICATION NUMBER: JP 260984/1994  
FILING DATE: 01-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP No. 6087146 yet received  
FILING DATE: 08-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: TSUSAKI=1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2899 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-863-010-4

Alignment Scores:  
Pred. No.: 1,66e-48 Length: 2889  
Score: 488.00 Matches: 147  
Percent Similarity: 42.39% Conservative: 87

Best Local Similarity: 26.63% Mismatches: 202  
Query Match: 14.32% Indels: 116  
DB: 3 Gaps: 22  
US-09-843-007A-2 (1-636) x US-08-863-010-4 (1-2889)  
Qy 118 GlyAspLeuLysGlyLeuLysAspLysIleProTyrPheGlnLeuGlyLeuThrTyr 137  
Db 82 GGGACTTTGAGGCGCTGAGCGGAGCTTCCCTACCTGAGGAGCTCGGGGTCAACACC 141  
Qy 138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAlaVal 157  
Db 142 CTCTGGCTCATGCTCTTCTTCAGTCCCTCC-----TTGAGGACCGAGGATCATC 195  
Qy 158 SerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIle 177  
Db 196 TCCGACTACTACCATCTCCCTCCCGCTCCAGCGACCTCGGAGCTTC-----ACCGTG 249  
Qy 178 AlaAlaLeuHisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSer 197  
Db 250 GACGAGCCCAACCGCGGGGATGAGGTGCTCATTTGAGCTCGTCCCTGAGACCACTCC 309  
Qy 198 AsnGluHisGluTTPAlaGlnArgCys---AlaAlaGlyAspProLeuPheAspAsnPhe 216  
Db 310 ATTGACCACTTGGTTCCAGGAGCGGAGCGGATAGCCCATCGCGGAC----- 363  
Qy 217 TyrTyrIlePheProAspArgA-gMetProAspGlnTyrAspArgThrLeuArgGluIle 236  
Db 364 TGGTACGTGTGAGCGGAC-----ACCCCGGAGAGTAC---AAGGGGGTCCGGGTCTATC 414  
Qy 237 PheProAspGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArgTTPVal----- 254  
Db 415 TTCAAGGAC-----TTTGAAACCTCCCACTGGACCTTTGAC 450  
Qy 255 -----TTPThrThrPheAsnSerPheGlnTTPAspLeuAsnTyr 267  
Db 451 CCGTGGCCCAAGGCTACTACTGGCACCGCTTCTACTGGCACAGCGCGACTCAACTGG 510  
Qy 268 SerAsnProTTPValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGly 287  
Db 511 GACAGCCCGAGGTGGAGAGGCGCATCCACAGCTCATGTCTTCTGGGCGGACTGGGG 570  
Qy 288 ValAspIleLeuArgMetAspAlaValAlaPheIleTTPLysGlnMetGlyThrSerCys 307  
Db 571 GTGACGCTTCCCGCTGGAGCGCATCCCTTACTCTACGAGCGGGAGCGGACCTCTGC 630  
Qy 308 GluAsnLeuProGlnAla---HisAlaLeuIleArgAlaPheAsnAlaValMetArgIle 326  
Db 631 GAGAACCTCCCGGAGACCATTTGAGCGGTGAGCGCTGAGAGGCGCTCGGAGGAGCGC 690  
Qy 327 AlaAlaProAlaValPhePheLysSerGluAlaIleValHisProAspGlnValValGln 346  
Db 691 TACGCGCCCGGAGAGATCTCTCTCGCGGAGCGCAACATGTGGCGGAGGACCTCTCCC 750  
Qy 347 TyrIleGlyGln---AspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeu 365  
Db 751 TACTTCGGGAGCGGGAGCGGGTCCACATGGCGCTTACCACTTCCCGCTGATGCCCGGATC 810  
Qy 366 TTPAsnThrLeuAlaThrArgGluValAsnLeuHisGlnAlaLeuThrTyrArgHis 385  
Db 811 TTCATGGCCCTTAAGCGGGAGGACCGGGGTCTCCCATTTGAACCATGCTCAAGAGCGGAG 870  
Qy 386 AsnLeuProGluHisThrAlaTTPValAsnTyrValArgSerHisAspAspIleGlyTyr 405  
Db 871 GGGATCCCGGAAACCGCGCGCTCTCTCTCGCGCAACACACGAGCTCACCGCTG 930  
Qy 406 ThrPheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPhe 425  
Db 931 GAGAGGCTCACGGAG-----GAGAGCGGGAGTTC 960  
Qy 426 LeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGln 445  
Db 961 ATGTACGAG-----GCCTACGCGCCCGGACCCCAAGTTCCGC 996

QY 446 TyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAlaAlaLeuValGlyLeu 465  
Db 997 ATCAAC  
QY 466 AlaGlnAspAspProHisAlaValAspArgIleLeuLeuLeuTyrSerIleAlaLeuSer 485  
Db 1033 CTCGGGGGCGACCGCAGG-----CGGTACGAGCTCTCCACCGCCCTCTCTCTCACC 1083  
QY 486 ThrGlyGlyLeuProLeuLeuLeuLeuLeuValGlyThrLeuAsnAspAsp--- 504  
Db 1084 CTAAAGGCGACGCCCATCTGCTACTACGGGGAGAGATCGGC---ATGGGGGCAACACCC 1140  
QY 505 -----AspTyrSerGlnAspSerAsnLys 512  
Db 1141 TTCCTCGGGGACCGAGCGGTGTGAGACCCCATCGAGTGTCTCCAGACCGCATCGTC 1200  
QY 513 SerAspAspSerArgTyrAlaHisArg-----ProArgTyrAsnGluAlaLeuTyr 529  
Db 1201 GCCTTCTCCCGCGCCCTACACAGCCCTCTCTCTCCCGCTGAGCGAGGGGCGCTAC 1260  
QY 530 -----AlaGlnArgAsnAspProSerThrAlaAlaGlyGln 541  
Db 1261 AGCTACCACTTCGTCAACGTGGAGGCGCCAGCGGGAACCCCTCCTCC----- 1308  
QY 542 IleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArg---PheAsp 560  
Db 1309 CTCCTGAGCTTCAACCGCGCTTCTCTCGCCCTGAGAACCGACCGCCAGATCTTCGCG 1368  
QY 561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArgAsn 580  
Db 1369 CGGGGAGCCTCACCTTCTCCCGTGGAGAACCGCGCGCTCTCTCTCTACCTGAGGAG 1428  
QY 581 Asn-----AlaLeuLeuAlaPheGlyAsnPheSerGlyThrProGlnThrValThr 597  
Db 1429 CACGAGGGGAGCGGCTCTCTCGTGTGGCCCACTCTCCCGCTACACCGCCCTTT--- 1485  
QY 598 AlaHisThrLeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyLysThrVal 617  
Db 1485 ----- 1485  
QY 618 SerLeuAsnGlnAspLeuThrLeuGlnProTyrGln 629  
Db 1486 -----GACCTTCCCTTGGAGGCTTACCA 1509

RESULT 10  
US-09-024-429-4  
Sequence 4, Application US/09024429  
Patent No. 6165768  
GENERAL INFORMATION:  
APPLICANT: TSUSAKI, Keiji  
APPLICANT: KUBOTA, Michio  
APPLICANT: SUGIMOTO, Toshiyuki  
TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR  
OPERATING SYSTEM: PC-DOS/MS-DOS  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/024,429  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/537,002

FILING DATE: 29-SEP-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 260984/1994  
FILING DATE: 01-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 255829/1995  
FILING DATE: 08-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: YUN, Allen C.  
REGISTRATION NUMBER: 37,971  
REFERENCE/DOCKET NUMBER: TSUSAKI=1B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2889 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-09-024-429-4  
Alignment Scores:  
Pred. No.: 1,668-48 Length: 2889  
Score: 488.00 Matches: 147  
Percent Similarity: 42.39% Conservative: 87  
Best Local Similarity: 26.63% Mismatches: 202  
Query Match: 14.32% Indels: 116  
DB: 3 Gaps: 22  
US-09-843-007A-2 (1-636) x US-09-024-429-4 (1-2889)  
QY 118 GlyAspLeuLysGlyLeuLysAspLysLeuProTyrPheGlnGluLeuGlyLeuThrTyr 137  
Db 82 GGGGATTTGGAGGCGCTGAGCGGAGCTTCCCTACCTGGAGGAGCTCGGGGTCAACACC 141  
QY 138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaVal 157  
Db 142 CTCGTGCTCATGCCCTTCTTCAGTCCCC-----TTGAGGAGCGACGCGGTACGATATC 195  
QY 158 SerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIle 177  
Db 196 TCGACTACTACAGATCTCTCCCTCCAGCGGAGCCCTGGAGGACTTC-----ACGGTG 249  
QY 178 AlaAlaLeuHisGluAlaGlyLysSerAlaValAspPheIlePheAsnHisThrSer 197  
Db 250 GACGAGGCGCCACGCGCGGAGTGAAGGTGATCATTTGAGCTCGTCTGAAACCCACCTCC 309  
QY 198 AsnGluHisGluTyrAlaGlnArgCys---AlaAlaGlyAspProLeuPheAspAsnPhe 216  
Db 310 ATTGACCACTTGGTTCAGGAGCGAGCGAGCGAGTACGCCCATCGGGAC----- 363  
QY 217 TyrTyrIlePheProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluIle 236  
Db 364 TGGTACGTGTGAGCGAC-----ACCCCGGAGAGTAC---AAGGGGGTCCGGGTGATC 414  
QY 237 PheProAspGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArgTyrVal----- 254  
Db 415 TTCAAGGAC-----TTTGAACCTCCAACTGGACCTTTGAC 450  
QY 255 -----TyrThrThrPheAsnSerPheGlnTyrPaspLeuAsnTyr 267  
Db 451 CCGGTGGCGAAGGCTTACTTGGACCGCTTCTACTGGACCGCTTCTACTGGACCGCTTCTACTGG 510  
QY 268 SerAsnProTyrValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGly 287  
Db 511 GACAGCGCCCGAGGTGGAGAGGCCATCCACCGAGGTGATGTCTTCTGGCGCGACCTGGGG 570  
QY 288 ValAspIleLeuArgMetAspAlaValAlaPheIleTyrLysGlnMetGlyThrSerCys 307  
Db 571 GTGGAGGCTTCCGCTGAGACGCCATCCCTACCTTACGAGCGGGAGGGGACCTCTCTGC 630



308 GluAsnLeuProGlnAla---HisAlaLeuIleArgAlaPheAsnAlaValMetArgIle 326  
|||  
631 GAGAACTCCCGAGACCAATTGAGCGGTGAGCGCTGAGGAGGCCCTGAGGAGCGC 690  
|||  
327 AlaAlaProAlaValPhePheIleValHisProAspGlnValGln 346  
|||  
691 TACGCCCGCGGGAAGATCTCTCTCGCGGAGGCAACATGTGCGCGAGGAGACCTCC 750  
|||  
347 TyrIleGlyGln---AspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeu 365  
|||  
751 TACTTCGGGAGCGGAGCGGCTCCACATGGCTCACTTCCCTGATGCCCCCGGATC 810  
|||  
366 TrpAsnThrLeuAlaThrArgGluValAsnLeuHisGlnAlaLeuThrTyrArgHis 385  
|||  
811 TTCTATGGCCCTAAGCGCGGAGGACCGGGCTCCATTGAACCACTCTCAAGGAGCGGAG 870  
|||  
386 AsnLeuProGluHisThrAlaTrpValAsnTyrValArgSerHisAspAspIleGlyTrp 405  
|||  
871 GGGATCCCGGAAACCGCCAGTGGCCCTTCTCCCGCAACCAACGACGAGCTCACCTG 930  
|||  
406 ThrPheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPhe 425  
|||  
931 GAGAAGGTCCACGAG---GAGGAGCGGAGTTC 960  
|||  
426 LeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGln 445  
|||  
961 ATGTACGAG---GCTACGCCCCCGACCCCAAGTTCCGC 996  
|||  
446 TyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeu 465  
|||  
997 ATCAAC---CTGGGATCCCGCGCGCTCATGCCCTC 1032  
|||  
466 AlaGlnAspAspProHisAlaValAspArgIleLeuLeuTyrSerIleAlaLeuSer 485  
|||  
1033 CTCGGGGCGACCGCAGG---CGGTACGAGCTCTCACCGCCCTCTCTCTCACC 1083  
|||  
486 ThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThrLeuAsnAsp--- 504  
|||  
1084 CTAAAGGCGACCCCATCTCTACTACGGGACGAGATCGGC---ATGGGGACACCC 1140  
|||  
505 ---AspTrpSerGlnAspSerAsnIlys 512  
|||  
1141 TTCCTCGGGACCGGAAACGGTGTACAGGACCCCATCATGCTGTCCTCCAAAGCCGATCGTC 1200  
|||  
513 SerAspAspSerArgTrpAlaHisArg---ProArgTyrAsnGluAlaLeuTyr 529  
|||  
1201 GCCTTCTCCCGCGCCCTTACCGCCCTTCTCTTCCCGCGGTGAGCGGGGCGCTTAC 1260  
|||  
530 ---AlaGlnArgAsnAspProSerThrAlaAlaGlyGln 541  
|||  
1261 AGCTACCACTTCGTCAACGTGAGGAGCCCGAGGAAACCCCTCC--- 1308  
|||  
542 IleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArg---PheAsp 560  
|||  
1309 CTCCTGAGCTTCAACCGCGCTTCTCTCGCCCTGAGGAACACGACGACCAAGATCTTCGGC 1368  
|||  
561 GlyGlyArgLeuValThrPheAsnThrAsnAsnIlysHisIleGlyTyrIleArgAsn 580  
|||  
1369 CGGGGAGCCTCACCTTCTCCCGTGGAGAACCGCGCTCTCTCGCTACTCTGAGGAG 1428  
|||  
581 Asn---AlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThr 597  
|||  
1429 CACGAGGGGAGCGGCTCTGTGTGGCCAACTCTCTCCGCTACACCCAGCGCTT--- 1485  
|||  
598 AlaHisThrLeuGlnAlaMetProPheIysAlaHisAspIleGlyGlyThrVal 617  
|||  
1485 --- 1485  
|||  
618 SerLeuAsnGlnAspLeuThrLeuGlnProTyrGln 629  
|||  
1486 ---GACCTCCCTTGGAGGCTTACCAA 1509  
|||

RESULT 11  
US-08-537-002A-5  
; Sequence 5, Application US/08537002A  
; Patent No. 5773282  
; GENERAL INFORMATION:  
; APPLICANT: TSUSAKI, Keiji  
; APPLICANT: KUBOTA, Michio  
; APPLICANT: SUGIMOTO, Toshiyuki  
; TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR  
; TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/537,002A  
; FILING DATE: 29-SEP-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 260984/1994  
; FILING DATE: 01-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP No. 5773282 yet received  
; FILING DATE: 08-SEP-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BROWDY, Roger L.  
; REGISTRATION NUMBER: 25,618  
; REFERENCE/DOCKET NUMBER: TSUSAKI=1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-528-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3600 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: DOUBLE  
; TOPOLOGY: linear  
; MOLECULE TYPE: genomic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Thermus aquaticus  
; INDIVIDUAL ISOLATE: ATCC 33923  
; FEATURE:  
; NAME/KEY: 5'UTR  
; LOCATION: 1..540  
; IDENTIFICATION METHOD: E  
; NAME/KEY: mat peptide  
; LOCATION: 541..3429  
; IDENTIFICATION METHOD: S  
; NAME/KEY: 3'UTR  
; LOCATION: 3430..3600  
; IDENTIFICATION METHOD: E  
US-08-537-002A-5  
Alignment Scores:  
Pred. No.: 2,41e-48 Length: 3600  
Score: 488.00 Matches: 147  
Percent Similarity: 42.39% Conservative: 87  
Best Local Similarity: 26.63% Mismatches: 202  
Query Match: 14.32% Indels: 116  
DB: 1 Gaps: 22  
US-09-843-007A-2 (1-636) x US-08-537-002A-5 (1-3600)

QY 118 GlyAspLeuGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyr 137  
DB 622 GGGAGCTTTGAGGGGCTTGAGGGGGAAGCTTCCCTACTGGAGAGCTCGGGGTCAACACC 681  
QY 138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaVal 157  
DB 682 CTCGTGCTCATGCTTCTTCAGTCCCC-----FTGAGGAGACAGCGGTACGATATC 735  
QY 158 SerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValle 177  
DB 736 TCCGACTACTACAGATCTCCCGCTCCAGCGGAGCCCTCGAGGACTTC-----ACCGTG 789  
QY 178 AlaAlaLeuHisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSer 197  
DB 790 GACGAGCCCAAGCGCGGGGATGAAGGTGATCATGTGAGCTGCTGTAACACACACCTCC 849  
QY 198 AsnGluHisGluTrpAlaGlnArgCys-----AlaAlaGlyAspProLeuPheAspHisPhe 216  
DB 850 ATGACACACCTTGGTTCCAGGAGCGAGAGCGAATAGCCCTATGCGGAC-----903  
QY 217 TyrTyrIlePheProAspArgMetProAspGlnTyrAspArgThrLeuArgGluIle 236  
DB 904 TGGTAGCTGTGAGCGAC-----ACCCCGAGAGATAC---AAGGGGTCCGGGTCACTC 954  
QY 237 PheProAspGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArgTrpVal-----254  
DB 955 TTCAGGAC-----TTTGAACCTCCCAACTGACCTTTGAC 990  
QY 255 -----TrpThrThrPheAsnSerPheGlnTrpAsnLeuAsnTyr 267  
DB 991 CCGGTGGCCAGGCTACTACTGGCACCGCTTCTACTGGCACAGCCCGACCTCAACTGG 1050  
QY 268 SerAsnProTrpValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGly 287  
DB 1051 GACAGCCCGAGGTGAGAGGCGATCCACAGGTGATGCTTCTTGGGCCGACCTGGGG 1110  
QY 288 ValAspIleLeuArgMetAspAlaValAlaPheIleTrpLysGlnMetGlyThrSerCys 307  
DB 1111 GTGGAGGCTTCCGCTGGACGCTCCCTCTACTCTACGAGGGAGGGACCTCTCTGC 1170  
QY 308 GluAsnLeuProGlnAla---HisAlaLeuIleArgAlaPheAsnAlaValMetArgIle 326  
DB 1171 GAGAACCTCCCGAGACCATTTAGGGGGGTGAAGCGCTGAGGAAGCGCTCGAGGAGGCG 1230  
QY 327 AlaAlaProAlaValPhePheLysSerGluAlaIleValHisProAspGlnValGln 346  
DB 1231 TAGCGCCCGGGAAGATCTCTCTCGCGGAGCGCAATGTGCGGAGGAGACCTCTCC 1290  
QY 347 TyrIleGlyGln---AspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeu 365  
DB 1291 TACTTCGGGAGCGGAGCGGGTCCACATGGCTACAACTTCCCTCGATGCCCGGATC 1350  
QY 366 TrpAsnThrIleAlaThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHis 385  
DB 1351 TTCTATGGCTTAAGCGGGAGGACCGGGGTCCCATTTGAACCATGCTCAAGGAGCGGAG 1410  
QY 386 AsnLeuProGluHisThrAlaTrpValAsnTyrValArgSerHisAspIleGlyTrp 405  
DB 1411 GGGATCCCGAARACCGCCAGTGGCGCTTCTCTCCGCAACACACAGCAGGTACCCCTG 1470  
QY 406 ThrPheAlaAspGluAspAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPhe 425  
DB 1471 GAGAGGTCAAGGAG-----GAGGAGCGGAGTTC 1500  
QY 426 LeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGln 445  
DB 1501 ATGTACGAG-----GCCATGCCCCCGACCCCAAGTTCGC 1536  
QY 446 TyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeu 465  
DB 1537 ATCAAC-----CTGGGATCGCGCGGCTCATGCCCTC 1572  
QY 466 AlaGlnAspProHisAlaValAspArgIleLysLeuLeuTyrSerIleAlaLeuSer 485

DB 1573 CTCGGGGCGACCGCAGG-----CGGTACGAGTCTCTACCGCCCTCTCTCTCACC 1623  
QY 486 ThrGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThrLeuAsnAspAsp---504  
DB 1624 CTAAAGGGCAGCGCCATGCTCTACTACGGGACGAGATCGGC-----ATGGGGACAAACCC 1680  
QY 505 -----AspTyrSerGlnAspSerAsnLys 512  
DB 1681 TTCCTGGGGACCGAAGCGGTGTCAGACCCCATCGAGTGTCTCCAGACCGCATCGTC 1740  
QY 513 SerAspAspSerArgTrpAlaHisArg-----ProArgTyrAsnGluAlaLeuTyr 529  
DB 1741 GCCTTCTCCGCGCCCT 1800  
QY 530 -----AlaGlnArgAsnAspProSerThrAlaAlaGlyGln 541  
DB 1801 AGTACCACTTCGTCAAGCTGAGCGCCCGAGCGGAAACCCCTCTCC-----1848  
QY 542 IleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArg---PheAsp 560  
DB 1849 CTCCTGAGCTTCAACCGCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1908  
QY 561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArgAsn 580  
DB 1909 CGGGGAGCGCT 1968  
QY 581 Asn-----AlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThr 597  
DB 1969 CACGAGGGGAGCGGTCTCTGTGTGGTCCCAACCTCTCTCTCTCTCTCTCTCTCTCT 2025  
QY 598 AlaHisThrLeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyThrVal 617  
DB 2025 -----2025  
QY 618 SerLeuAsnGlnAspLeuThrLeuGlnProTyrGln 629  
DB 2026 -----GACCTCCCTCTGGAGCGCTTACCA 2049

RESULT 12  
US-08-863-010-5  
; Sequence 5, Application US/08863010  
; Patent No. 6087146  
; GENERAL INFORMATION:  
; APPLICANT: TSUSAKI, Keiji  
; APPLICANT: KOBOTA, Michio  
; APPLICANT: SUGIMOTO, Toshiyuki  
; TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR  
; TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/863,010  
; FILING DATE: 01-SEP-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/537,002  
; FILING DATE: 29-SEP-1995  
; APPLICATION NUMBER: JP 260984/1994  
; FILING DATE: 01-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP No. 6087146 yet received



Qy 618 SerLeuAsnGlnAspLeuThrLeuGlnProTyrGln 629  
Db 2026 -----GACCTCCCTTGGAGGCTACCAA 2049

RESULT 13  
US-09-024-429-5  
Sequence 5, Application US/09024429  
Patent No. 6165788  
GENERAL INFORMATION:  
APPLICANT: TSUSAKI, Keiji  
APPLICANT: KUBOTA, Michio  
APPLICANT: SUGIMOTO, Toshiyuki  
TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME FOR  
TITLE OF INVENTION: CONVERTING MALTOSE INTO TREHALOSE  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESS: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/024, 429  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/537,002  
FILING DATE: 29-SEP-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 260984/1994  
FILING DATE: 01-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 255829/1995  
FILING DATE: 08-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: YUN, Allen C.  
REGISTRATION NUMBER: 37,971  
REFERENCE/DOCKET NUMBER: TSUSAKI-1B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3600 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: DOUBLE  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Thermus aquaticus  
INDIVIDUAL ISOLATE: ATCC 33923  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 1..540  
IDENTIFICATION METHOD: E  
NAME/KEY: mat peptide  
LOCATION: 541..3429  
IDENTIFICATION METHOD: S  
NAME/KEY: 3'UTR  
LOCATION: 3430..3600  
IDENTIFICATION METHOD: E  
US-09-024-429-5  
Alignment Scores:

Pred. No.: 2,41e-48 Length: 3600  
Score: 488.00 Matches: 147  
Percent Similarity: 42.39% Conservative: 87  
Best Local Similarity: 26.63% Mismatches: 202  
Query Match: 14.32% Indels: 116  
DB: 3 Gaps: 22  
US-09-843-007A-2 (1-636) x US-09-024-429-5 (1-3600)  
Qy 118 GlyAspLeuLysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyr 137  
Db 622 GGGGACTTTGAGGGCTTGGAGGAGAGCTTCCCTTACCTGGAGGAGCTCGGGACCAACACC 681  
Qy 138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaVal 157  
Db 682 CTCGGGCTCATGCCCTTCTCCAGTCCCC-----TTGAGGGACGAGCGGTACGATC 735  
Qy 158 SerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValLe 177  
Db 736 TCCGACTACTACAGATCTCCCGCTCCAGGGACCTGGAGGACTTC-----ACCGTG 789  
Qy 178 AlaAlaLeuHisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSer 197  
Db 790 GACGAGGCCACGGCCGGGGATGAAGTGATCATTTAGGTCGCTCGTGAACCACTCC 849  
Qy 198 AsnGluHisGluTrpAlaGlnArgCys---AlaAlaGlyAspProLeuPheAspAsnPhe 216  
Db 850 ATTGACCACCTTGGTTCAGAGGGCGAGGAAGCCGAATAGCCCATTCGCGGAC----- 903  
Qy 217 TyrTyrIlePheProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluLe 236  
Db 904 TGGTACGTGTGGAGCGAC-----ACCCCGAGGAGTAC---AAGCGGTCTCGGGGTATC 954  
Qy 237 PheProAspGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArgTrpVal----- 254  
Db 955 TTCAAGGAC-----TTTGAACCTTCAACTGGACCTTTGAC 990  
Qy 255 -----TrpThrPheAsnSerPheGlnTrpAspLeuAsnTyr 267  
Db 991 CCCGTGCCAAGGCTTACTCTGGACCGCTTACTGGACACAGCCGACCTCACTTGG 1050  
Qy 268 SerAsnProTrpValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGly 287  
Db 1051 GACAGCCCGAGGTGGAGAGGCCATCCACGAGTCATGTCTTCTGGGCGGACCTGGGG 1110  
Qy 288 ValAspIleLeuArgMetAspAlaValAlaPheIleTrpLysGlnMetGlyThrSerCys 307  
Db 1111 GTGACCGCTTCCGCTGGAGCGCATCCCTTACGAGCGGAGGAGGACCTCTGTC 1170  
Qy 308 GluAsnLeuProGlnAla---HisAlaLeuIleArgAlaPheAsnAlaValMetArgile 326  
Db 1171 GAGAACCTCCCGAGACCATTTGAGCGGCTGAGCGGCTGAGGAGGCGCTCGAGGAGCGC 1230  
Qy 327 AlaAlaProAlaValPhePheLysSerGluAlaIleValHisProAspGlnValGln 346  
Db 1231 TACGGCCCGGGAAGATCTCTCTCGCGAGGCCAACATGTGGCGGAGGAGGACCTCCCC 1290  
Qy 347 TyrIleGlyGln---AspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeu 365  
Db 1291 TACTTGGGAGCGGGAGCGGGTCCATGCGCTTACAACTTCCCGCTGATCCCCGGATC 1350  
Qy 366 TrpAsnThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHis 385  
Db 1351 TTCATGGCCCTTAGCGCGGAGGACCGGGGTCTCCATTGAAACCATGCTCAAGAGGCGGAG 1410  
Qy 386 AsnLeuProGluHisThrAlaTrpValAsnTyrValArgSerHisAspAspIleGlyTrp 405  
Db 1411 GGGATCCCCGAAACCGCCCGAGTGGGCGCTCTTCTCCGCAACACGAGGAGCTCACCTG 1470  
Qy 406 ThrPheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPhe 425  
Db 1471 GAGAGGTCACGGAG-----GAGGAGCGGGAGTTC 1500

QY 426 LeuAsnArgPheValAsnArgPheAspGlySerPheAlaValGlyValPropheGln 445  
Db 1501 ATCTAGCAG-----GCCATGCCCGCCGACCCCAAGTTCCCG 1536  
QY 446 TyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAlaValGlyLeu 465  
Db 1537 ATCAAC-----CTGGGATCCCGCGCCCTCAGCCCTC 1572  
QY 466 AlaGlnAspProHisAlaValAspArgGlyLeuLeuTyrSerIleAlaLeuSer 485  
Db 1573 CTGGGGGGGACCGCAGG-----CGTACGAGCTCTCAGCCCTCTCTCTCACC 1623  
QY 486 ThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThrLeuAsnAspAsp 504  
Db 1624 CTAAGGGGACCGCCCATCGTCTACTAGGGGAGAGATCGGC---ATGGGGGACACCC 1680  
QY 505 -----AspTyrSerGlnAspSerAsnLys 512  
Db 1681 TTCTCTGGGACCGGAACGGTGTTCAGGACCCCATCGAGTGGTCCCAAGACCGCATCTC 1740  
QY 513 SerAspAspSerArgTrpAlaHisArg-----ProArgTyrAsnGluAlaLeuTyr 529  
Db 1741 GCCTTCTCCCGCCCGCCCTACCAACCGCTCTCTCTCCCGCCCGGAGCGGCGCTAC 1800  
QY 530 -----AlaGlnArgAsnAspProSerThrAlaAlaGlyGln 541  
Db 1801 AGTACACCTCTCTCACTGAGGAGGCGCGGAGAAACCCCATCTCC----- 1848  
QY 542 IleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArg-----PheAsp 560  
Db 1849 CTCTGAGCTTCAACCGCGCTCTCTCGCCCTCAGGAACCGACGACCGCAAGATCTTCGG 1908  
QY 561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArgAsn 580  
Db 1909 CGGGGAGGCGCTACCTCTCTCCCGGAGAACCGCGGCTCTCTCTCTCTCTCTCTCTCT 1968  
QY 581 Asn-----AlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThr 597  
Db 1969 CAGAGGGGAGCGGCT 2025  
QY 598 AlaHisThrLeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyLysThrVal 617  
Db 2025 -----GACCTCCCTTGGAGGCTACCAA 2049  
QY 618 SerLeuAsnGlnAspLeuThrLeuGlnProTyrGln 629  
Db 2026 -----GACCTCCCTTGGAGGCTACCAA 2049

RESULT 14  
US-09-107-532A-2079  
; Sequence 2079, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 2079:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1728 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHEetical: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...1728  
SEQUENCE DESCRIPTION: SEQ ID NO: 2079:  
US-09-107-532A-2079

Alignment Scores:  
Pred. No.: 9,6e-39 Length: 1728  
Score: 404.50 Matches: 156  
Percent Similarity: 40.97% Conservative: 105  
Best Local Similarity: 24.43% Mismatches: 277  
Query Match: 11.8% Indels: 99  
DB: 4 Gaps: 28

US-09-843-007A-2 (1-636) x US-09-107-532A-2079 (1-1728)

QY 34 SerGluAspTyrArgGlnPheSerArgMetAspThrHisPheProLysLeuMetAsn 53  
Db 1 TCATACTCTGGGACGCTCCCAAAAAGAAAGGGGAAT-----CCGATGTATACAAA 54  
QY 54 GluLeuAspSerValTyrGlyAsnAsnGluAlaLeuLeuProMetLeuGlnMetLeuLeu 73  
Db 55 CTTTTCACAAAGCATTTATCGGACTCAGAT---TTGTCTGAGAAACCGCCCTCTATGATT 111  
QY 74 AlaGlnAlaTyrGlnSerTyrSerGlnArgAsnSerSerLeuLysAspIleAla 93  
Db 112 GAACAAAAGTCAGTGTGTTTATCAGCAGAGAAAGGCCGAGGCAAACTG----- 162  
QY 94 ArgGluAsnAsnProAspTyrIleLeuSerAsnLysGlnValGlyValCysTyrVal 113  
Db 163 -----CATTTG-----GACGAGAAGATATTTTCTGATTACTTATGGC 201  
QY 114 AspLeuPheAlaGlyAspLeuLysGlyLeuLysAspLysIlePro----- 128  
Db 202 GATCAATTT-----TTTGAAGAGCAGCAACCACTTTGCGTACATTT 243  
QY 129 -----TyrPheGlnGluLeuGlyLeuThrTyr-----LeuHisLeuMetProLeu 143  
Db 244 AGAAATTTTATCAAAAGATTTTGGCAGATACATTTAGATTGTGTACATTTCTTGCAATTT 303  
QY 144 PheLysCysProGluGlyLysSerAspGlyTyrAlaValSerSerTyrArgAspVal 163  
Db 304 TTCCCATAT-----ACATCTGATGATGTTTCTCCGTTGTAATTTATCAGCAATC 354  
QY 164 AsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAlaLeuHisGluAla 183  
Db 355 AACGACCGTTTGGGACATGGGAAGATATAGAAATG-----AGAAGC 399  
QY 184 GlyIleSerAlaValValAspPheIlePheAsnHisThrSerAsnGluHisGluTrpAla 203  
Db 400 GAAACACAGGTTAATGTTGCTGCTTCTGCTCATCATATGCTGCGGAAAGTACTTGTTT 459



```

403  TATATTGGGCAGATCTCTAAA---CCGGAT----- 429
QY  ProAspGlnHisProGlyGlyPheSerGlnLeuGluAspGly----- 251
Db  430  ---GGATCTGAACCTTAATACTGGGAAGATCTTTTAATGGTTCACTGGGAGTTTGAC 486
QY  252  -----ArgTrpValTrpThrPheAsnSerPheGlnTrpAspLeuAsnTyr 267
Db  487  GAATCGACTAAGCAATACTATTTCCATTTATTATAGCAAAAGCAGCCAGATTATTAATTGG 546
QY  268  SerAsnProTrpValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGly 287
Db  547  GAAATCCAGATGTAGACAAGCTGTGTTGAATGATGAATGGTGGTGGTTGAAAAGGT 606
QY  288  ValAspIleLeuArgMetAspAlaValAlaPheIleTrpLysGlnMetGlyThrSerCys 307
Db  607  ATTGACGGATTAGAGTTGATGTCATTTACTATTAAGAAAGATTTTGAAGCAGAGAT 666
QY  308  GluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPheAsnAlaValMetAlaGlyAla 327
Db  667  TTACCTGTACTGTGTCGCAAAATTTGCTCCAGCATTTGATGTAGATATGAAT----- 720
QY  328  AlaProAlaVal-----PhePheLysSerGluAlaIle----- 338
Db  721  CAGCCAGGAATACAAGAATGCTCCAGAAATGAAGATAAATCGTTAAGTCGGTATGAC 780
QY  339  -----ValHisProAspGlnValValGlnTyrIleGly 349
Db  781  ATTATGACTGTAGCGAGGCTAATGGTGTACTCTTAATGATGCTGAAGATGGGTAGGA 840
QY  350  GlnAspGluCysGlnIleGlyTyrAsnProLeu-----GlnMetAlaLeuLeuTyr 365
Db  841  GAAGAAAATGGGAAA-----TTTAATATGATATTCCAGTTTGAACATCTTGGTTATGG 894
QY  367  AsnThrLeuAlaThrArg---GluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHis 385
Db  895  AGTACTGCGATACGAAATTCGATGTTAAATCCCTATAAACAAAGCTTAAATCGTTGGCAA 954
QY  386  AsnLeuProGluHisThrAlaTrp---ValAsnTyrValArgSerHisAspAsp----- 402
Db  955  AAGCAACTAGAAATGTAGTTGGATGCTTTATTATTCGAAACCATTGATCAACACGT 1014
QY  403  -----IleGlyTrpThrPheAlaAspGluAspAlaIleTyrLeuGlyIleSerGlyTyr 420
Db  1015  CGTGTTCACACTGG-----GGTGATGATAAAATTTATGTTGATGAATCAAGCAACT 1065
QY  421  AspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArg 440
Db  1066  AGTCAC-----GCTACTGCCCTACTTTTACAA-----CAG 1095
QY  441  GlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAla 460
Db  1096  GGCACACCTTTTATTTACCAAGTCAAGAA----- 1125
QY  461  AlaLeuValGlyLeuAlaGlnAspAspProHisAlaValAsp-----Arg 475
Db  1126  -----ATAGGTATGACTAATTAATTCATTGTAAGCACTTGAAGTTTCAACGATGTCGA 1179
QY  476  IleLysLeuLeuTyrSerIleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGly 495
Db  1180  GTGAATACTGATATCAATAGTCAAAAGAGGTGGA---GATGTCAATCAATTAATACTA 1236
QY  496  AspGluValGlyThrLeuAsnAspAspAsp-----TyrSerGlnAsp 509
Db  1237  GATAAATATAAATGGAAAACCGAGACCAATCCAAAGACTCCAATCCAATGGAATTAATCT 1296
QY  510  SerAsnLys-----SerAspAspSerArgTTPAlaHis---ArgProArgTyrAsnGlu 526
Db  1297  ATCAATGCTGGATTCTACTGCTAGGCAATGTTTCATGTAAACCTAACTATACAGAA 1356
QY  527  AlaLeuTyrAlaGlnArgAsnAspProSerThrAlaAlaGlyGlnIleTyrGlnGlyLeu 546

```

```

Db  1357  ATTAATGTTAAACAAACAATAAATGATTAAGTTTTCGATACTTTCTTATTATAAAGCGTTA 1416
QY  547  ArgHisMetIleAlaValArgGlnSerAsnProArgPheAspGlyGlyArgLeuValThr 566
Db  1417  -----ATTCAACTAAAAAACTGATTTGATTTGATTTTACACCTACGTTAAGTTTAATATG 1467
QY  567  PheAsnThrAsnAsnLysHisIleIleGlyTyrIleArg-----AsnAsnAlaLeu 583
Db  1468  GTCGATGCTGAAAATAAAGCAGGTTTTTTCATATACACGCACATTTAAAAACAATACTGTA 1527
QY  584  LeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThrLeuGlnAla 603
Db  1528  TTAATTTGAGCCCAATCTCAAAATGAAGTATCAGAACTAAAC----- 1569
QY  604  MetProPheLysAlaHisAspLeuIleGlyGlyLysThrValSerLeuAsn----- 620
Db  1570  CTACCTTTTGAATTAGAT-----ATTTCATCTGTAGATATAAAATTCATAATTATCAC 1623
QY  621  ---GlnAspLeuThrLeu-----GlnProTyrGlnValMetTrpLeuGluIle 635
Db  1624  TTAATGATATAAAATTTAGACCAATATAAAACCTTATGAATCATTCGTCGTTGAATA 1680

```

Search completed: November 8, 2003, 23:30:29  
Job time : 2977 secs



```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 8, 2003, 21:58:19 ; Search time 486 Seconds
        (without alignments)
        4175.821 Million cell updates/sec

Title: US-09-843-007A-2
Perfect score: 3408
Sequence: 1 MLPTQGVGLILYKTRIL.....VSLNQDLTLQPYQMMLA 636

Scoring table:
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.0
Dloep 6.0 , Delext 7.0

Searched: 2141354 seqs, 1595478879 residues

Total number of hits satisfying chosen parameters: 4282708

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xl0
-Q=/cgn1/USPFO_Spool_P/US09843007/runat_07112003_170422_25467/app_query.fasta_1.775
-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALLCN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pt0 -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09843007.ecgn1.1.107@runat_07112003_170422_25467
-NCPU=6 -ICPU=3 -NO MAP -LARGUEURY NEG SCORES=0 -WAIT -DSBLOCK=100
-LONGLOP -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09E_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result      Query
No.         Score Length DB ID      Description
-----
1 3408 100.0 2914 9 US-09-843-007-1 Sequence 1, Appli

```

ALIGNMENTS

RESULT 1

```

US-09-843-007-1
; Sequence 1, Application US/09843007
; Patent No. US20020092040A1
; GENERAL INFORMATION:
; APPLICANT: Kossmann, Jens
;             Buttcher, Volker
;             Welsh, Thomas
; TITLE OF INVENTION: DNA SEQUENCES ENCODING ENZYMES CAPABLE
;                   OF FACILITATING THE SYNTHESIS OF LINEAR
;                   ALPHA-1,4 GLUCANS IN PLANTS, FUNGI AND
;                   MICROORGANISMS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: James P. Haley, Jr., c/o Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)

```

```

Sequence 7764, Ap
Sequence 2790, Ap
Sequence 2527, Ap
Sequence 1, Appli
Sequence 1, Appli
Sequence 7374, Ap
Sequence 53, Appli
Sequence 153, App
Sequence 8494, Ap
Sequence 4455, Ap
Sequence 13, Appli
Sequence 191, App
Sequence 171, App
Sequence 167, App
Sequence 159, App
Sequence 173, App
Sequence 655, App
Sequence 157, App
Sequence 1713, Ap
Sequence 1, Appli
Sequence 185, App
Sequence 189, App
Sequence 161, App
Sequence 183, App
Sequence 657, App
Sequence 11, Appli
Sequence 5957, Ap
Sequence 15, Appli
Sequence 6367, Ap
Sequence 95, Appli
Sequence 4, Appli
Sequence 97, Appli
Sequence 307, App
Sequence 99, Appli
Sequence 644, App
Sequence 7, Appli
Sequence 8, Appli
Sequence 9889, Ap
Sequence 21784, A
Sequence 56, Appli
Sequence 1, Appli
Sequence 2, Appli
Sequence 2, Appli

```

```

541.5 15.9 3303 9 US-09-815-242-7764
529.5 15.5 1716 14 US-10-156-761-2790
529.5 15.5 1794 10 US-09-738-626-2527
529.5 15.5 3309400 10 US-09-738-626-1
529.5 15.5 9025608 14 US-10-156-761-1
511.5 15.0 1749 14 US-10-156-761-7374
338 9.9 16592 8 US-08-781-986A-53
331 9.7 1620 14 US-10-081-872-153
328.5 9.6 1650 9 US-09-815-242-8494
323.5 9.5 1638 9 US-09-815-242-4455
313.5 9.2 1782 14 US-10-061-269-13
300.5 8.8 1596 14 US-10-081-872-191
300 8.8 1431 14 US-10-081-872-171
300 8.8 1596 14 US-10-081-872-167
299.5 8.8 1587 14 US-10-081-872-159
299 8.8 1596 14 US-10-081-872-173
298 8.7 1683 10 US-09-974-300-655
295.5 8.7 1596 14 US-10-081-872-157
295.5 8.7 1599 14 US-10-156-761-1713
294.5 8.7 9025608 14 US-10-156-761-1
292.5 8.6 1572 14 US-10-081-872-185
292.5 8.6 1596 14 US-10-081-872-189
292 8.6 1623 14 US-10-081-872-161
291.5 8.6 1596 14 US-10-081-872-183
287 8.4 1689 10 US-09-974-300-657
283.5 8.3 1794 14 US-10-061-269-11
283 8.3 1848 14 US-10-156-761-5957
280 8.2 1704 14 US-10-061-269-15
276.5 8.1 1656 9 US-09-815-242-6367
274.5 8.1 1524 14 US-10-081-872-95
32 274.5 8.1 1803 12 US-10-223-277-4
33 274.5 8.1 1803 12 US-10-223-277-5
34 274.5 8.1 2931 14 US-10-081-872-97
35 273 8.0 4888 10 US-09-070-927A-307
36 271.5 8.0 1536 14 US-10-081-872-99
37 271.5 8.0 1703 10 US-09-974-300-644
38 265.5 7.8 1659 12 US-10-223-277-7
39 265.5 7.8 1659 12 US-10-223-277-8
40 263.5 7.7 1653 9 US-09-815-242-9889
41 262.5 7.7 3354 12 US-09-814-353-21784
42 262.5 7.7 30246 8 US-08-781-986A-56
43 260 7.6 1656 12 US-10-223-277-1
44 260 7.6 1656 12 US-10-223-277-2
45 253.5 7.4 2863 12 US-10-050-763-2

```



QY 501 LeuAspAspAspTirpSerGlnAspSerAsnLysSerAspSerArgTrpAlaHis 520  
 DB 2457 CTCATAGCAGGAGCTGTGCGAAGACCAATAGAGGACGACAGCGTTGGCGGCAC 2516  
 QY 521 ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAppProSerThrAlaAlaGly 540  
 DB 2517 COTCGCGCTACACGAGCCCTGTACGCGCAACGCAATCCGTCGACGCGCGG 2576  
 QY 541 GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp 560  
 DB 2577 CAAATCTATCAGGCTTGGCCATATGATTCGCGTCCGCAAGCAATCCGCGCTTCGAC 2636  
 QY 561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArgAsn 580  
 DB 2637 GCGCGCAGCTGGTTACATTTCAACCAACCAACAGCACATCATCGCTACATCCGCAAC 2696  
 QY 581 AsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThr 600  
 DB 2697 AATGCGCTTTTGGCAATTCGGTAATTCAGCGCAATATCCGCAACCGTTACCGCGCATACC 2756  
 QY 601 LeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyLysThrValSerLeuAsn 620  
 DB 2757 CTGCAAGCCATGCCCTTCAGGCGGACGACCTCATCGTGGCAAACTGTGAGCTGGAT 2816  
 QY 621 GlnAspLeuThrLeuGlnProTyrGlnValMetTrpLeuGluIleAla 636  
 DB 2817 CAGGATTTGACCTTCAGCCCTATCAGGTCTATGCTGGCTCGAAATCGCC 2864

## RESULT 2

US-09-815-242-7764  
 ; Sequence 7764, Application US/09815242  
 ; Patent No. US20020061569A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari L.  
 ; APPLICANT: Zyskind, Judith W.  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John D.  
 ; APPLICANT: Carr, Grant J.  
 ; APPLICANT: Yamamoto, Robert T.  
 ; APPLICANT: Xu, H. Howard  
 ; TITLE OF INVENTION: Identification of Essential Genes in  
 ; FILE REFERENCE: ELITRA 011A  
 ; CURRENT APPLICATION NUMBER: US/09/815,242  
 ; CURRENT FILING DATE: 2001-03-21  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; NUMBER OF SEQ ID NOS: 14110  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 7764  
 ; LENGTH: 3303  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)...(3303)  
 US-09-815-242-7764

Alignment Scores: 7.86e-56 Length: 3303  
 Pred. No.:

Score: 541.50 Matches: 165  
 Percent Similarity: 41.60% Conservative: 90  
 Best Local Similarity: 26.92% Mismatches: 223  
 Query Match: 15.89% Indels: 135  
 DB: 9 Gaps: 25  
 US-09-843-007A-2 (1-636) x US-09-815-242-7764 (1-3303)  
 QY 96 AsnAsnProAspTrpIleLeuSerAsnLysGlnValGlyGlyValCysTyr----- 112  
 DB 31 AACACCCGCAATGGTACAAGGAC-----GGGTGATCTACAGGTCCAC 75  
 QY 113 ValAspLeuPhe-----AlaGlyAspLeuLysGlyLeuLysAsp 125  
 DB 76 GTGAAATCCCTTCTACGAGCGCAACACGATGGCATCGGCACTTCGCGGCTGATCGAG 135  
 QY 126 LysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeuMetProLeuPheLys 145  
 DB 136 AAGCTGACTACATCGCGGACCTCGCGGTGAACACTCTCTGGCTGTGCTCCCTTCTAC--- 192  
 QY 146 CysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyrArgAspValAsnPro 165  
 DB 193 ---CCGTGCGCACCGCCGAGCGGCTACGACATCGCCCGTACCGGAGGCCATCGCGGCTCTG 309  
 QY 166 AlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAlaLeuHisGluAlaGlyIle 185  
 DB 250 GACTACGCGAGCTCGCGGACGCGCGGTTCATCGCGGAGGCCATCGCGGCTCTG 309  
 QY 186 SerAlaValValAspPheIlePheAsnHisThrSerAsnGluHisGluTrp----- 202  
 DB 310 CGGTGATTTACCGAGCTGGTGATCAACATACCTTCGACGACGATCCCTGGTTCATCGT 369  
 QY 203 AlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePheProAsp 222  
 DB 370 GCCCGCACGCAAGAGGATCGCGCGCGCGGAC-----TACTACGCTCTGGTGGAC 423  
 QY 223 ArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAsnGlnHisPro 242  
 DB 424 -----AGCGACGAGAAATACCGAGGTACG---CGGATCATCTTCATCGAC--- 465  
 QY 243 GlyGlyPheSerGlnLeuGluAspGlyArgTrpValTrpThr----- 256  
 DB 466 -----ACCGAGCAGTCCCACTGCGACCTGGGACCGCGTAGCCCAACAGTAC 510  
 QY 257 -----ThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsnProTrpValPhe 273  
 DB 511 TACTGGCACCGCTTCTATTCCACCGAGCGGACTGAACTTCGACCAACCGCGAGTCCGT 570  
 QY 274 ArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMet 293  
 DB 571 CGCGAGTGTCTCGGGTGATGCGCTACTGCTGACATGCGGCGTTCGACGCGCTGGCGCTG 630  
 QY 294 AspAlaValAlaPheIleTyrPheLysGlnMetGlyThrSerCysGluAsnLeuProGlnAla 313  
 DB 631 GACCGGATTCCTTACCTGATCGAAGCGAGCGACCGAGCGAGAGAACCTTCGCGGAGACC 690  
 QY 314 HisAlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaProAlaValPhePhe 333  
 DB 691 CACAGGTCTCAAGCGGCTCCGCGCGGAGTCCGCGGCGGACTATCCCGACCGCATGCTG 750  
 QY 334 LysSerGluAlaIleValHisProAspGlnValValGlnTyrIleGly----- 349  
 DB 751 CTGCGCGAGGCCAATCAGTGGCGCGGAAGACACCGCGCGGTACTTCGCGCGGAGGATGCG 810  
 QY 350 -----GlnAspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTrpAsn 367  
 DB 811 GCGAGGCGGAGCAATGCGACATGGCGCTTCCACTTCGCGTGTATGCGCGCATGTACATG 870  
 QY 368 ThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeu 387  
 DB 871 GCCATCGCGCAGGAGGATCGCTATCCGATCACCAGACATCTCGCGCAGACCCCGGACATC 930  
 QY 388 ProGluHisThrAlaTrpValAsnTyrValArgSerHisAspAspIleGlyTrpThrPhe 407

```

Db 931 CCGGCCAATGGCAATGGCGGATCTTCTGCGCAACACGACGAGCTCCCTCGGAGATG 990
QY 408 AlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsn 427
Db 991 GTACCGAC
QY 428 ArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsn 447
Db 1018 -----TGGAAC 1023
QY 448 ProSerThrGlyAspCysArgValSer-----GlyThrAlaAlaAlaValGly 464
Db 1024 CACTATGCCCGCGACCGCGCGCGCCCTCAACCTGGGCGATCCGCGCGCGCTCGCGCG 1083
QY 465 LeuAlaGlnAspAspProHisAlaValAspArgIleLysLeuLeuTyrSerIleAlaLeu 484
Db 1084 CTGGTGGAGCGTGCACCGCGG-----CGCATCGAGCTGCTGCACAGCCTGTGTG 1134
QY 485 SerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr----- 500
Db 1135 TCGATGCCCGGCGACCGCGGACCTGTACTACGCGGACGAGATCGGCATGCGGCGACACATC 1194
QY 501 ---LeuAsnAspAspAsp-----TrpSerGlnAspSerAsn--- 511
Db 1195 TACCTCGCGGACCGCGGCGGCGGCGGACCGCGATGCACTGCTGCGTGGACCGCAACGCG 1254
QY 512 -----LysSerAspAspSerArgTrpAlaHisArgProArgTyrAsnGluAlaLeu 528
Db 1255 GCGTTCTCCCGCGCGACCGCGGAGAGAGCTGGTG---CTGCGCGCGATTCCTGACCGCGTG 1311
QY 529 Tyr-----AlaGlnArgAsnAspProSerThrAlaAlaGly 540
Db 1312 TACGGCTACACGAGCATCAAGTCGAGGCGGCGGCGGCGGACCGCATTCG----- 1362
QY 541 GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp 560
Db 1363 ---CTGCTCAACTGGATGCGCGCTGCTGCGCGTGGCGGACGAGCATCTCTGCGTGGCGGCGGCGG 1419
QY 561 GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIle----- 578
Db 1420 CCGCGGAGCTGAAGTGTGCGCGCGCGGACGAGCATCTCTGCGTGGCGGCGGCGGCGGCGG 1479
QY 579 -----ArgAsnAsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrPro 593
Db 1480 TACGCGGAGCGGAGCGGAGGACGAGCATCTCTGCGTGGCGGCGGCGGCGGCGGCGGCGG 1539
QY 594 GlnThrValThrAlaHisThrLeuGlnAlaMetProPheLysAlaHis----- 609
Db 1540 CAGCGCGTG-----GAACGTGACCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1581
QY 610 -----AspLeuIleGlyLysThrVal-----SerLeuAsnGlnAspLeu 623
Db 1582 CCGGTGGAGATGATCGCGGCGATGCTGTTCCCGCGGATCGCGGCGGCGGCGGCGGCGGCGG 1641
QY 624 ThrLeuGlnProTyrGlnValMetTrpLeuGluIleAla 636
Db 1642 ACCCTCGCGCGCTACGCGCTTCTACTGCTTCTACTGCTTCTACTGCTTCTACTGCTTCT 1680
```

## RESULT 3

```

US-10-156-761-2790
; Sequence 2790, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
```

```

; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2790
; LENGTH: 1716
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1716)
US-10-156-761-2790
Alignment Scores:
Pred. No.: 8,49e-55 Length: 1716
Score: 529.50 Matches: 166
Percent Similarity: 41.31% Conservative: 86
Best Local Similarity: 27.21% Mismatches: 233
Query Match: 15.54% Indels: 125
DB: 14 Gaps: 23
```

```

US-09-843-007A-2 (1-636) x US-10-156-761-2790 (1-1716)
QY 93 AlaArgGluAsnAsnProAspTrpIle----- 101
Db 46 GCCAAGGACCCGATCCGGAGTGTTCAAACGCCCGCTCTTCTACGAGGTCTCGTCGC 105
QY 102 ---LeuSerAsnLysGlnValGlyIleValCysTyrValAspLeuPheAlaGlyAspLeu 120
Db 106 TCCTTCCAGGACACGACGCGGCGGCTGTC-----GGCGACCTG 144
QY 121 LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeu 140
Db 145 AAGGCTCTGACCGCCAGCTGACTATCTGCGTGGCTGGCGGCGGAGTGTCTGTGGCTG 204
QY 141 MetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAlaValSerSerTyr 160
Db 205 CCGCGCTTCTTCAAGTCCCGC-----CTGCGCGACGCGGCTACGAGCTCTCGGACTAC 258
QY 161 ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAlaLeu 180
Db 259 ACCGCGTGTGCGCGAGTTCGTCGAGTTCGTCGAGTTCGTCGAGTTCGTCGAGCGGCC 318
QY 181 HisGluAlaGlyIleSerAlaValAspPheIlePheAsnHisThrSerAsnGluHis 200
Db 319 CACCAGCGCGCATCGCGGTGATCATCGACTTCGTGATGAACACACACGACGACCTGCAC 378
QY 201 GluTrpAlaGln-----ArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyr 218
Db 379 CCGTGGTTCAGGAGTCCCGACGACCCCGACGCGCCCTTACGCGAC-----TACTAC 432
QY 219 IlePheProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePhePro 238
Db 433 GTGTGGCGCGAC-----GACGACAGCAGTACGAGGAGCGCGCGGATCATCTTCGTC 483
QY 239 AspGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArgTrpVal----- 254
Db 484 GAC-----ACCGAGGCGCTCCAACTGACCTTCGACCCCGGTC 519
QY 255 -----TrpThrThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsn 269
Db 520 CGCAAGCAGTACTACTGGCAGCGCTCTTCTCCACGACGCGGACCTCACTACGAGAAC 579
QY 270 ProTrpValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAsp 289
Db 580 CCGCGCGTGCAGGAGAGATCGTCTCGCGCTCTGGCTCTGGCTCGACCTCGGCATCGAC 639
QY 290 IleLeuArgMetAspAlaValAlaPheIleTrpLysGlnMetGlyThrSerCysGluAsn 309
Db 640 GCGTTCGCGCTGACGCGGTGCGGTACTGTGTACGAGGAGGAGGACCACTCGCGAAC 699
```

Qy	310	LeuProGlnAlaHisAlaLeuLeuArgAlaPheAsnAlaValMetArgIleAlaAlaPro	329
Db	700	CTGCGCGGCGACGACGAGTTCCTGAAGCGGGTGGCGAAGAGATCGACGCACTATCCG	759
Qy	330	AlaValPhePheLysSerGluAlaIleValHisProAspGlnValValGlnTyrIleGly	349
Db	760	GACACGGTGTCTGCGGAGCGAACCAGTGGCGGAGGACGTGGTGCATCTTCGGC	819
Qy	350	Gln-----AspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeu	364
Db	820	GACTTCCTCCGCGGCGGACGACGATCGGCATGTCGCTTCATTTCCCGGTCATCCGCGG	879
Qy	365	LeuTrpAsnThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArg	384
Db	880	ATCTTCATGGCGGTGGCGGTGAGTGGCGGTATCCGGTTCGGAGATCTTGGCGAAGCG	939
Qy	385	HisAsnLeuProGluHisThrAlaTrpValAsnTyrValArgSerHisAspAspIleGly	404
Db	940	COGCGGATCCGTCGAGTGCAGTGGCGGATCTTCTGCGCAACACGACGACTCACC	999
Qy	405	TrpThrPhe---AlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArg	423
Db	1000	CTGGAGATGTGTACCGAGGAGACGCGACTACATGTGG-----	1038
Qy	424	GlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValPro	443
Db	1039	-----GCGAGTACGCCAAGATCCGCGG	1062
Qy	444	PheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAlaLeuVal	463
Db	1063	ATCGCGGCCAAC-----ATCGGCATCCCGCGCGCTCTGGCG	1098
Qy	464	GlyLeuAlaGlnAspAspProHisAlaValAspArgIleLysLeuLeuTyrSerIleAla	483
Db	1099	CCGCTGTGTGACAAACGACCGC-----AACCAGATCGAGCTGTTCACCGCGCTGCTG	1149
Qy	484	LeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr-----	500
Db	1150	CTGTGCTGCCCGCTCCCGATCTCTATACGCGCAGAGATGGGATGGGGGACAA	1209
Qy	501	-----LeuAsnAspAspAsp-----TrpSerGlnAspSerAsn	511
Db	1210	ATCTGGCTCGGTGACCGGACGCGGTGGCGACTCCGATCGAGTCGACGCGACCGCAAC	1269
Qy	512	-----LysSerAspAspSerArgTrpAlaHisArgProArgTyrAsnGluAlaLeu	528
Db	1270	GCGGGTTTCGTCTCGACCCCGGGCGTCTGTATCGCCACGATCATGGATCCCGGT	1329
Qy	529	Tyr-----AlaGlnArgAsnAspProSerThrAlaAlaGly	540
Db	1330	TACGGTACCAGTCCAGAACGTGGAGCGTCGATGTCGTGGCGCTCTCTCG-----	1380
Qy	541	GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp	560
Db	1381	---CTGTGCACTGGACCGCGCGGATGATCGAGATCCGTATAGCAGAACCCCGCGTCTCGGC	1437
Qy	561	GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArg---	579
Db	1438	CTCGGCTGTACACCGAACTCCAGTCTCTGAAACCCCGCGCTCTCGGTCTCTCGGGAG	1497
Qy	580	-----AsnAsnAlaLeuLeuAlaPheGlyAsnPheSerGlu	591
Db	1498	GCCCTCTGACCGGGGGAACGGGGACGACCTGTGTGTGTGTGTCGTCACACTCTCTCCGG	1557
Qy	592	TyrProGlnThrValThrAlaHisThrLeuGlnAlaMetProPheLysAlaHis-----	609
Db	1558	TTCCGCGAGCCC---ACGGAGCTGGATCTGGGGGGTTCACGGCGCGT---CATCCGTC	1611
Qy	610	AspLeuIleGlyLysThrVal-----SerLeuAsnGlnAspLeuThrLeu	625
Db	1612	GAGCTGATCGCGGTGTCCGCTTCCCGGCCATCGGGGAACCTCCGCTATCTGTCACCGCTG	1671

```

Cy      626 GlnProTyrglnValMetTrpLeuGluIle 635
Db      1672 GCAGGCCACGGCTTCTACTGTTCCGGCTC 1701
      ::::
      |||
      :
RESULT 4
US-09-738-626-2527
; Sequence 2527, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCES: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 2527
; LENGTH: 1794
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-2527

Alignment Scores:
Pred. No.: 9,11e-55 Length: 1794
Score: 529.50 Matches: 153
Percent Similarity: 46.32% Conservative: 105
Best Local Similarity: 27.47% Mismatches: 218
Query Match: 15.54% Indels: 81
DB: 10 Gaps: 22

US-09-843-007A-2 [1-636] X US-09-738-626-2527 [1-1794]

```

Alignment Scores:			
Pred. No.:	9.11e-55	Length:	1794
Score:	529.50	Matches:	153
Percent Similarity:	46.32%	Conservative:	105
Best Local Similarity:	27.47%	Mismatches:	218
Query Match:	15.54%	Indels:	81
DB:	10	Gaps:	22
US-09-843-007A-2 (1-636) x US-09-738-626-2527 (1-1794)			
Qy	118	GlyAspLeuLysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyr	137
Db	223	GGATCGTTGAAGCCCTGACCGAARAACCTGGATTACATCCAGTGGCTCGCGCTGATATGC	282
Qy	138	LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaVal	157
Db	283	ATTGGATCCACCGTTTATGATTCGCCA-----CTGCGCGACGGCGGTACCATATC	336
Qy	158	SerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIle	177
Db	337	CGCAACTTCGGTGAATTCCTGCCGAATTCGGCCAGCTCGATTCGATTCGTGGAATCTGTT	396
Qy	178	AlaAlaLeuHisGluAlaGlyIleSerAlaValAlaAspPheIlePheAsnHisThrSer	197
Db	397	GACCACGCCACCGCGCTGGCTCGGTGTATCACCGACTGGTCATGAATCACACCTCC	456
Qy	198	AsnGluHisGluTrpAlaGlnArgCysAlaAlaGlyAspProLeu-----PheAspAsn	215
Db	457	GACCAGCACGCATGGTTCCAAGAA---TCCCGCGCGACCCACACCGGCCCTTACGGAGATC	513
Qy	216	PheTyrTrpIlePheProAspArgMetProAspGlnTyrAspArgThrLeuArgGlu	235
Db	514	TTCATGTGTGGACCGATGAT-----CCCACCTGTATACACGAAGCC---CGCATC	561
Qy	236	IlePheProAspGlnHisProGlyGlyPheSer---GlnLeuGluAspGlyArgTrpVal	254

562 ATCTTTTATAGATACAGAGAAATCCAACTGAGCACTATGATCCGGTGGCTGGCAGTACTTC 621  
QY TTPThrThr-PheAsnSerPheGlnTTPAspLeuAsnTyrSerAsnProTTPValPheArg 274  
Db TGGACCGCTTCTTCCACACACAGACCTCACTACGACAAACCCCGCAGTCCAGAG 681  
QY AlaMetAlaGlyClnMetLeuPheLeuAlaAsnLeuGlyValAspLeuArgMetAsp 294  
Db GCCATGCTAGATGCTTGGCTTCTGGCTGACCTGGGCTTGGCTTGGCTTGGCTTGGCT 741  
QY AlaValAlaPheIleTTPAspLeuMetGlyThrSerCysGlnAsnLeuProGlnAlaHis 314  
Db GCGTTCCTTATCTTTTGGACCGGAGGAGGACCAACCGCGGAAACCTCAAGAGAACCCAC 801  
QY AlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaProAlaValPhePheLys 334  
Db GATTTCTCAAACTGTGCTCTGTCACTGAGAGAGAAATACCCCGCGGAATCTCTGCTC 861  
QY SerGlnAlaIleValHisProAspGlnValIleGlnTyrIleGln----- 350  
Db GCGAGGACCAACATGGCCCAAGATGGTGGATATCTTCGGTGAAGAAAGCAAGGC 921  
QY AspGlnCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTTPAsnThrLeuAla 370  
Db GATGAATGCCATGCTGCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCT 981  
QY ThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHis 390  
Db CAAGTTCCAGCCCGCCGATGAGATCTCTGCCCAACACCCCGGAGATCTCCAGACT 1041  
QY ThrAlaTTPValAsnTyrValArgSerHisAspIleGlyTTPThrPhe-----AlaAsp 409  
Db GCCCATGGGTATTTTCTGCTGATATCATGATGAGCTACCTCCCTGAAATGGTCTCCGAT 1101  
QY GluAspAlaIleTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsn----- 427  
Db GAGAACCGCAGCTACATG-----TACTCCCAATTCGCTCCGAACT 1143  
QY ArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsn 447  
Db CGCATGCGCCCAAC----- 1158  
QY ProSerThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAlaGln 467  
Db GTAGGAATCCGAGCGGCGCTTTCCCGCAGCTCTGAA 1194  
QY AspAspProHisAlaValAspArgIleValLeuLeuTyrSerIleAlaLeuSerThrGly 487  
Db AACCGACCGC-----AACCGACCGC-----AACCGACCGC-----AACCGACCGC 1245  
QY GlyLeuProLeuIleTyrLeuGlyAspGlnValGlyThr-----LeuAsn 502  
Db GGCTACCGCTGTGTATTAGCGTGTATGAAATGGCATGGCGGACCAATATCTGGCTCCAC 1305  
QY AspAspAsp-----TTPSerClnAspSerAsn----- 511  
Db GACCGCAGCGAGTGGCAGCCCGCCATGCTGGTCCAAACCGCAGCGGCTGTCTCTCC 1365  
QY LysSerAspAspSerArgTTPAlaHisArgProArgTyrAsnGlnAlaLeu-----TyrAla 530  
Db AAGCTGATCTGAAGCGCTGTACTTCCAGCGATCCAAATATGATCAATACGGCTACGCC 1425  
QY GlnArgAsnAspProSer-----ThrAlaAlaGlyClnIleTyrGlnGlyLeuArgHis 548  
Db CAAGTAAACGTGGAAGCCCAACTCAACCGCGGAAACCTCCCTGTGGCGCTCCGAAAC 1485  
QY MetIleAlaValArgGlnSerAsnProArgPheAspGlyGlyArgLeuValThrPheAsn 568  
Db CAATCTTATCCGACAGCAGTACCGCATTTGGTCCCGGACCTACCGTGAAGTCTCC 1545  
QY ThrAsnAsnLysHisIleIleGlyTyrIleArgAsnAsn-----AlaLeuLeuAla 585  
Db TCCACCAATGAGTCAAGTGTGACATTTTACGACACACAGAGGCGCAACCATTTGTGT 1605

QY 586 PheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThrLeuGlnAlaMetPro 605  
Db 1606 GTCAACACATGAGCAATATCTCTCAGGCAGTC-----TCGCTTGAATTCGGTGAA 1656  
QY 606 PheLysAlaHis-----AspLeuIleGlyGlyLys-----ThrValSerLeu 619  
Db 1657 TTTGACGACACACACCCCTCGAGAGATGTCGGCGGCGAGCTGTCTCCCTACCATTCGTGA 1716  
QY 620 AsnGln-----AspLeuThrLeuGlnProTyrGlnValMetTTPLeuGluIle 635  
Db 1717 CGGAGTGGATTTGTCACCTTTAGCCCTCAGCGATTTCTTCTGTTTGATCTC 1767  
RESULT 5  
US-09-738-626-1  
Sequence 1, Application US/09738626  
Publication No. US20020197605A1  
GENERAL INFORMATION:  
APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI  
APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO  
APPLICANT: OCHIAI, KEIKO  
APPLICANT: YOKOI, HARUHIKO  
APPLICANT: TATEISHI, NAOKO  
APPLICANT: SENOH, AKIHIRO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738,626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 1  
LENGTH: 3309400  
TYPE: DNA  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-1  
Alignment Scores:  
Pred. No.: 1-5e-49 Length: 3309400  
Score: 529.50 Matches: 153  
Percent Similarity: 46.33% Conservative: 105  
Best Local Similarity: 27.47% Mismatches: 218  
Query Match: 15.54% Indels: 81  
DB: 22 Gaps: 22  
US-09-843-007A-2 (1-636) x US-09-738-626-1 (1-3309400)  
QY 118 GlyAspLeuGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyr 137  
Db 2438335 GATCGTTGAAAGGCTGACCGAAACCTGATATCATCTCGCTCGCGGTGATTCG 2438394  
QY 138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaVal 157  
Db 2438395 ATTTCGATCCACCGTTTTATGATTTCCCA-----CTCGCGGACGGCGTTACGATATC 2438448  
QY 158 SerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIle 177  
Db 2438449 CGCAATCTCCGTAATCTCTGCCGAATTCGCGACCGCTGATGACTTCGTGGAACCTCGT 2438508  
QY 178 AlaAlaLeuHisGlnAlaGlyIleSerAlaValAlaAspPheIlePheAsnHisThrSer 197  
Db 2438509 GACCACGCCACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2438568  
QY 198 AsnGluHisGluTTPAlaGlnArgCysAlaAlaGlyAspProLeu-----PheAspAsn 215





DB 3441682 AAGGGCTGACCGCAAGCTGGACTATCTGAGTGGTGGCGGTGGAGTGGCTGTGGCTG 3441741  
QY 141 MetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyr 160  
DB 3441742 CCGCGTCTTCAAGTCCCC-----CTGCGGACGCGGCTAGAGCTCTCCGACTAC 3441795  
QY 161 ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAlaLeu 180  
DB 3441796 ACCGGGTGCTGCCCGAGTTCGGTGCAGCTCTCGTCCGAGTTCGTGCGACGCGGCC 3441855  
QY 181 HisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSerAsnGluHis 200  
DB 3441856 CACACGCGCGGTGCGGTGATCATCGACTTCGTGATGAACCAACACCGGACGACTGCAC 3441915  
QY 201 GluTrpAlaGln-----ArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyr 218  
DB 3441916 CCGTGTCTCCAGGAGTCCCGGACCAACCCGACGCGCCCTACGCGGAC-----TACTAC 3441969  
QY 219 IlePheProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePhePro 238  
DB 3441970 GTTGGGCGGAC-----GACGACAGCAGTACACGAGGAGCCCGGATCATCTTGTCT 3442020  
QY 239 AspGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArgTyrVal-----254  
DB 3442021 GAC-----ACCGAGGCTCCAACTGGACCTTCGACCGGTC 3442056  
QY 255 -----TrpThrPheAsnSerPheGlnTyrAspLeuAsnTyrSerAsn 269  
DB 3442057 CGCAAGCAGTACTACTGCGACCGCTTCTTCCACGCGCGGACCTCAACTAGAGAAC 3442116  
QY 270 ProTrpValPheArgAlaMetAlaGlyGlnMetLeuPheLeuAlaAsnLeuGlyValAsp 289  
DB 3442117 CCGCGGTGCGAGGAGATCGTCTCCGCGCTCGGTTCTGGCTCGACCTCGCATCGAC 3442176  
QY 290 IleLeuArgMetAspAlaValAlaPheIleTrpLysGlnMetGlyThrSerCysGluAsn 309  
DB 3442177 GGTTCGCGCTCGACGCGGTGCGTACCTGTACACGAGGAGGACCAACTGCGAGAAC 3442236  
QY 310 LeuProGlnAlaHisAlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaPro 329  
DB 3442237 CTGCGCGGACGACAGAGTCTGGAAGCGGTGCGCAGGAGATCGACACGACTATCCG 3442296  
QY 330 AlaValPhePheLysSerGluAlaIleValHisProAspGlnValValGlnTyrIleGly 349  
DB 3442297 GACACGGTGTCTCGCGGAGCGGACCAAGTGGCGGAGGACGTGTGCTGACTACTTCGCG 3442356  
QY 350 Gln-----AspGluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeu 364  
DB 3442357 GACTTCCGCTCGCGCGGCGGACGAGTCCACATCGGTTCCTATTTCCCGGTATCCCGCGG 3442416  
QY 365 LeuTrpAsnThrLeuAlaThrArgGluValAlaAsnLeuHisGlnAlaLeuThrTyrArg 384  
DB 3442417 ATCTTCATGGCGTCCGCGTGGTGGTATCCGTTGCGGATCTCGGAGATCTCGGAGACG 3442476  
QY 385 HisAsnLeuProGluHisThrAlaTrpValAsnTyrValArgSerHisAspIleGly 404  
DB 3442477 CCGCGATCCCGTCCGAGTGGCGGTCATCTTCTCGGCAACCAACGACGAGTGCACC 3442536  
QY 405 TrpThrPhe---AlaAspGluAspAlaValTyrLeuGlyIleSerGlyTyrAspHisArg 423  
DB 3442537 CTGGAGATGGTCAACGACGAGGACCGGACTACATGTG-----3442575  
QY 424 GlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValPro 443  
DB 3442576 -----GCGGATGACGCCAAGATCCGCGG 3442599  
QY 444 PheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAlaLeuVal 463  
DB 3442600 ATGCGGCGCAAC-----ATCGGATCCGCGCGGTCTGCG 3442635  
QY 464 GlyLeuAlaGlnAspAspProHisAlaValAspArgIleLysLeuLeuTyrSerIleAla 483  
DB 3442636 CGCTGCTGGACACGACGCG-----AACGAGATGAGCTGTTCACGCGGTGCTG 3442686

## RESULT 7

US-10-156-761-7374  
; Sequence 7374, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 7374  
; LENGTH: 1749  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1749)  
US-10-156-761-7374

Alignment Scores:

Pred. No.: 1.49e-52 Length: 1749

Score:	511.50	Matches:	163
Percent Similarity:	40.98%	Conservative:	87
Best Local Similarity:	26.72%	Mismatches:	235
Query Match:	15.01%	Indels:	125
DB:	14	Gaps:	23

US-09-843-007A-2 (1-636) x US-10-156-761-7374 (1-1749)

93	AlaArgGluAsnAsnProAspTrpIle	101
Qy		
Db	GCACAGGACCGGATCCGAGTGGTTCAACAGCGCCGGCTTCTACAGAGGTCTCTGGTCCGC	105
Qy	102 ---LeuSerAsnLysGlnValGlyGlyValCysTyrValAspLeuPheAlaGlyAspLeu	120
Db	106 TCCTTCCAGGACAGCAACGGCACGGTGTG	144
Qy	121 LysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrtTyrLeuHisLeu	140
Db	145 AAGGGCTTACCCCAAGCTGACTACTCTGCAGTGGTGGCGTGGAGTGCCTGTGGCTG	204
Qy	141 MetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAlaValSerSerTyr	160
Db	205 CGCGCGTCTTCAAGTCCCC	258
Qy	161 ArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaLeu	180
Db	259 ACCGGTGCTCCCGAGTTCGGTCACTGGCCGACTTCGTGAGTTCGTGGAGCGCGCC	318
Qy	181 HisGluAlaGlyIleSerAlaValAspPheIlePheAsnHisThrSerAsnGluHis	200
Db	319 CACCAGCGCGCATGGCGTGATCATCGACTTCGTGATGACACACACAGTGCATGCAT	378
Qy	201 GluTrpAlaGln	218
Db	379 CGGTGGTTCNCGAGTCCCGCAGCAACCCGACGGCCCTACGGCGAC	432
Qy	219 IlePheProAspArgMetProAspGlnTyrAspArgThrLeuArgGluIlePhePro	238
Db	433 GTGTGGCCGAC	483
Qy	239 AspGlnHisProGlyPheSerGlnLeuGluAspGlyArgTrpVal	254
Db	484 GAC	519
Qy	255	269
Db	520 CGCAACGACTACTCTGGCACCGCTTCTCTCCACAGCCGACCTCAACTACGAGGT	579
Qy	270 ProTrpValPheArgAlaMetAlaGlyIleMetLeuPheLeuAlaAsnLeuGlyValAsp	289
Db	580 GCGCGCTCAGAGGAGATCTGGCGGCGCTGGGTCTGGCTCGATCTGGGCATCGAC	639
Qy	290 IleLeuArgMetAspAlaValAlaPheIleTrpLysGlnMetGlyThrSerCysGluAsn	309
Db	640 GGCTTCAGGCTGCAGCCGCTCCCTACCTGTACACGAAGAGGGGACAACTGCGAGAAC	699
Qy	310 LeuProGlnAlaHisAlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaPro	329
Db	700 CTGCGCGGACGACGAGTCTCTGAAGCGGGTGGCGAAGAGATTCGACAGCATATCCG	759
Qy	330 AlaValPhePheLysSerGluAlaIleValHisProAspGlnValValGlnTyrIleGly	349
Db	760 GACACGGTCTGCTCGGGAGCGGAACCGATGGCGGGAGGAGTGGTTCGACTACTTCGGC	819
Qy	350 Gln	364
Db	820 GACTTCCCTCGGGCGGACGAGTGCACATGGCGTTCATTTCCATTTCCCGCGTCATCGCGCG	879
Qy	365 LeuTrpAsnThrLeuAlaThrArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArg	384
Db	880 ATCTTCATCGCGTGGCGGTGAGTCCGGGTATCCGGTGTCCGAGATCTCTGGCGAAGACG	939
Qy	385 HisAsnLeuProGluHisThrAlaTrpValAsnTyrValArgSerHisAspAspIleGly	404

Db	940	CGCGCATCCGTCGAGCTCCGATGGGGCATCTTCTCGCGAACACGACGAGCTGACC	999
Qy	405	TrpThrPhe---AlaAspGluAspAlaAlaItyrLeuGlyIleSerGlyTyrAspHisArg	423
Db	1000	CTCGAGATGCTACCGACGAGGAACGCGACTACATGTGG---	1038
Qy	424	GlnPheIeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyValpro	443
Db	1039	-----CGCGAGTACGCCACAGGATCCGCGG	1062
Qy	444	PheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAlaLeuVal	463
Db	1063	ATCGGGCCCAAC-----ATCGCATCCGCCGCGTCTCGCG	1098
Qy	464	GlyLeuAlaGlnAspAspProHisAlaValAspArgIleIysLeuLeuTyrSerIleAla	483
Db	1099	CGGCTGTGTGACACAGCACCGC-----AACCATGACGCTGTTCACCGCGCTCTGTG	1149
Qy	484	LeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlyThr---	500
Db	1150	CTGTGCTGCCGCGCTCGCCGATCTCTACTACGGCAGAGATCGCGATGGGGACAAC	1209
Qy	501	-----LeuAsnAspAspAsp-----TyrSerGlnAspSerAsn	511
Db	1210	ATCTGGCTCGGTACCGGGACCGCGTCCGACCGCGATGCGAGTGACCGGACCGGCAAC	1259
Qy	512	-----LysSerAspAspSerArgTyrAlaHisArgProArgTyrAsnGluAlaLeu	528
Db	1270	GCGSGTTTCTCGTCTCGACCGGGCGCTGTGTATCGCCACGATCATGTGATCGGTC	1329
Qy	529	Tyr-----AlaGlnArgAsnAspProSerThrAlaAlaGly	540
Db	1330	TACGGGTACAGGTACGAACGTGGAGCGTGCATGTCTCGCGCTCTCG-----	1380
Qy	541	GlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAsp	560
Db	1381	---CTGCTGCATCGACCGCGCGATCATCGAGATCGGTAAAGCAGAACCCGGCGTTCGCG	1437
Qy	561	GlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyrIleArg---	579
Db	1438	CTCGGCTCGTACACCGAACTCCAGTCTCTGAACCGGCGCTCTCTCGGTCGCGGAG	1497
Qy	580	-----AsnAsnAlaLeuLeuAlaPheGlyAsnPheSerGlu	591
Db	1498	GCCCCCTCGACCGGGGGAACGGGACGACCTGTGTGTGTGCGGTGCACACTTCTCCCGG	1557
Qy	592	TyrProGlnThrValThrAlaHisThrLeuGlnAlaMetProPheLysAlaHis-----	609
Db	1558	TTCGCGCAGGCC---ACGGAGCTGGATCTCGGGCGGTTCAGCGGCGCGT---CATCCGCGTC	1611
Qy	610	AspLeuIleGlyGlyLysThrVal-----SerLeuAsnGlnAspLeuThrLeu	625
Db	1612	GAGCTGATCGCGGTGTCCGCTCCCGGCCATCGGGAACTCCCGTATCTGCTCACCCTG	1671
Qy	626	GlnProTyrGlnValMetTyrLeuGluIle	635
Db	1672	GCGGGCCACGGCTTCTACTGTTCGCGCTC	1701

## RESULT 8

US-08-781-986A-53/c

US-08-731-300A-357C  
: Sequence 53, Application US/08781986A

; Publication No. US20030054436A1

GENERAL INFORMATION:

APPLICANT: Charles Kunsch

; TITLE OF INVENTION: St

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue  
CITY: Miami, Florida

CITY: Rockville  
COUNTY: Montgomery

STATE: Maryland  
COUNTRY: USA

; ; COUNTRY: USA

ZIP: 20850  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
 COMPUTER: HP Vectra 486/33  
 OPERATING SYSTEM: MSDOS version 6.2  
 SOFTWARE: ASCII Text  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/781,986A  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Benson, Bob  
 REGISTRATION NUMBER: 30,446  
 REFERENCE/DOCKET NUMBER: PB248PP  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (301) 309-8504  
 TELEFAX: (301) 309-8512  
 INFORMATION FOR SEQ ID NO: 53:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 16592 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear

US-08-781-986A-53

Alignment Scores:  
 Pred. No.: 1,736-29 Length: 16592  
 Score: 338.00 Matches: 158  
 Percent Similarity: 36.36% Conservative: 106  
 Best Local Similarity: 21.76% Mismatches: 230  
 Query Match: 9.92% Indels: 232  
 DB: 8 Gaps: 34

US-09-843-007a-2 (1-636) x US-08-781-986A-53 (1-16592)

Qy 6 GlnGlnValGly---LeuIleLeuGlnTyrLeuLysThrArg-----IleLeuAsp 21  
 Db 4830 CAACAGCGTGGTGCAGCAATTTAAATATATTGAAACAAACAGCAGGATTTATTGAA 4771  
 Qy 22 IleTyrThrProGluGlnArgAlaGlyIleGluLysSerGluAspTyrArgGlnPheSer 41  
 Db 4770 CTCGTCATTATAGATACAGATTAATAAATAAGAAATCAACACAGCGATAGCAAGGAGT 4711  
 Qy 42 ArgArgMetAspThrHisPheProLysLeuMetAsnGluLeuAspSerValTyrGlyAsn 61  
 Db 4710 TTTAAATGAATAACCAATGGTGAA-----4684  
 Qy 62 AsnGluAlaLeuLeuProMetLeuGluMetLeuLeuAlaGlnAlaTyrGlnSerTyrSer 81  
 Db 4683 ---GAAGCAGTA-----GCATATCAAGTATATCCA 4657  
 Qy 82 GlnArgAsnSerSerLeuLysAspIleAspIleAlaArgGluAsnProtsPyrIle 101  
 Db 4656 AGA-----AGTTTAAATGAT-----AGTAATCACAGTGGTATT 4624  
 Qy 102 LeuSerAsnLysGlnValGlyGlyValCysTyrValAspLeuPheAlaGlyAspLeuLys 121  
 Db 4623 -----GGGGAATTACCT 4612  
 Qy 122 GlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLeuMet 141  
 Db 4611 GGAATGATTGATTAATGAGCTACTTAAAGATTTCGGTATCGATGCTATTGGCTCAGT 4552  
 Qy 142 ProLeuPheLysCysProGluGlyLysSerAspGlyTyrAlaValSerTyrArg 161  
 Db 4551 CCAATGTTTAAATACCT-----AATGATGACAAATGGTTATGATATTAGTACTACCAA 4498  
 Qy 162 AspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaLeuHis 181  
 Db 4497 GAGATTATGGATGAATTGGACGATGGGAAGACTTTGATCGTTTATTAAAGGTGTTTAT 4438

Qy 182 GluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSerAsnGluHisGlu 201  
 Db 4437 GATAGAGCGATGAAGCTTTATTAGATTGTTGTAATCATACATCTGATGAACTCT 4378  
 Qy 202 Trp-----AlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrIle 219  
 Db 4377 TGGTTTATAGATCCAAATCTAGTAAAGACAAATCCCAACAGTGAT-----TGGTACATT 4324  
 Qy 220 PheProAspArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAsp 239  
 Db 4323 TGGCAAGATCCAAAG---CCAGAT-----GGC 4300  
 Qy 240 GlnHisProGlyGlyPheSerGlnLeuGluAspGly-----251  
 Db 4299 TCTGAACCTTAACAACTGGGGAAGTATATTAAATGATCTCATGCGAATATGCTAAT 4240  
 Qy 252 -----ArgTrpValTrpThrPheAsnSerPheGlnTrpAspLeuAspTyrSerAsn 269  
 Db 4239 ACTGACCAATATTATTTCATTTTTCAGTAAACCAACCTGATTGAAATGGGGTAAT 4180  
 Qy 270 ProTrpValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaLeuGlyValAsp 289  
 Db 4179 CCGAAGTTAGATGCTGTATTGAAATGATGACTGGTGGTTTGTATGAAGGCAATTGAT 4120  
 Qy 290 IleLeuArgMetAspAlaValAlaPheIleTrpLysGlnMetGlyThrSerCysGluAsn 309  
 Db 4119 CGATTTAGATGATGCAATTACGCATATTAAAGAGACCTTTGAAGCGGTGACTTACCT 4060  
 Qy 310 LeuProGlnAlaHisAlaLeuIleArgAlaPheAsnAlaValMetArg-----325  
 Db 4059 GTACCTGAGGTAAACATATGCCCGACGATTGTTGATGATGATGATGATGATGATGAT 4000  
 Qy 326 -----IleAlaAlaProAlaValPhePheLys 334  
 Db 3999 CAAACTTGGTTACAAGAGATGAAGATCGCTCATTAAGTAAGTATGATGATGATGATGAT 3940  
 Qy 335 SerGluAlaIle---ValHisProAspGlnValValGlnTyrIleGlyGlnAspGluCys 353  
 Db 3939 GGTGAACGAATGGTGAACCCCTGATGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTG 3880  
 Qy 354 GlnIleGlyTyrAsnProLeu-----GlnMetAlaLeuLeuTrpAsnThrLeuAla 370  
 Db 3879 AAA-----TTTAAATGATATTCCCAATTGAAACATTTGGGACTGTGGAATAGTGGTAT 3826  
 Qy 371 ThrArg---GluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGlu 389  
 Db 3825 TCTCAGCTTGTATGATAATTCGTATAATCTGTATTAAATAGATGGCAAAACAACTTGA 3766  
 Qy 390 HisThrAlaTrp---ValAsnTyrValArgSerHisAspAspIleGly-----TrpThr 406  
 Db 3765 AATAAGCTTGGATGGTGGTTGTTTATTGAAATCAACGCCAACCCAGCGTGTATCGACG 3706  
 Qy 407 PheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeu 426  
 Db 3705 TGGGGTCAAGATGACAAAGTATTGG-----TATGAA-----3676  
 Qy 427 AsnArgPheValAsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyr 446  
 Db 3676 -----3676  
 Qy 447 AsnProSerThrGlyAspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAla 466  
 Db 3675 -----TCAGCAACAAAGTCATGCACTGTTATTCTTCTG 3643  
 Qy 467 GlnAspAspProHisAlaValAspArgIleLysLeuLeuTyrSerIleAlaLeuSerThr 486  
 Db 3642 CAACAA-----3637  
 Qy 487 GlyGlyLeuProIleTyrLeuGlyAspGluValGly-----499  
 Db 3636 ---GGTACGCCATTCAATTATCAAGGTCAAGAAATGGTATGACCAATATTCATTTCAA 3580

```
QY 500 -----ThrLeuAsnAsp-----AspAspTTPSer----- 507
Db 3579 AGTATTGAAACGTTTAAACGATGCTGCTTTAAATGACTATCAATAGTGAAGCTCAA 3520
QY 508 -----GlnAspSerAsnLysSerAspAspSer 516
Db 3519 GGTGGAGATGTAGACGCTTTACTTCCGAAATATAAAGATGAGAACCGA---GATAATTCT 3463
QY 517 Arg-----TrpAlaHis----- 520
Db 3462 CGCACACCAATGCAATGGATGATACGTTAAATGGAGGATTTACAAATGGTGAACCGTGG 3403
QY 521 -----ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspProSerThr 537
Db 3402 TTCACAGTAATCCGAATTATAAAACTATCAATGTGTCACAACTAGTAAGATGAGCAT 3343
QY 538 AlaAlaGlyGlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnPro 557
Db 3342 TCAGTATTCAATTTTAAAGATTTA-----ATTCAATTAAAGAAAGTCTTAATCAT 3292
QY 558 ArgPheAspGlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyr 577
Db 3291 GTATACGTATATGTCATTTGATTTAGTACATGCTGAAATTCACAAATTTTTCGCTAC 3232
QY 578 IleArg-----AsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrProGln 594
Db 3231 ACAGAAACATTAATGAAGCAAGTCTTATTGTAGGTAATCTTACTAACCCAGCA 3172
QY 595 ThrValThrAlaHisThrLeuGlnAlaMetProPheLysAlaHisAspLeuIleGly 614
Db 3171 GAATTAAT-----GTACCAAT-----GATTAAGCCATGA 3139
QY 615 LysThrValSerLeuAsnGlnAspLeu-----ThrLeuGlnProTyrGln 629
Db 3138 GAACTAAAGCTATTAAATATGATGCAAGTAAATTAACAGTTACGTCCTATATGAA 3079
QY 630 ValMetTrpLeuGluIle 635
Db 3078 GCATGTGTTATCGAACTA 3061
```

## RESULT 9

```
US-10-081-872-153
; Sequence 153, Application US/10081872
; Publication No. US20030125534N1
; GENERAL INFORMATION:
; APPLICANT: Callen, Walter
; APPLICANT: Richardson, Toby
; APPLICANT: Frey, Gerhard
; APPLICANT: Short, Jay M.
; APPLICANT: Mathur, Eric J.
; APPLICANT: Gray, Kevin A.
; APPLICANT: Kerovuo, Janne S.
; APPLICANT: Slupska, Malgorzata
; TITLE OF INVENTION: ENZYMES HAVING ALPHA AMYLASE ACTIVITY
; TITLE OF INVENTION: AND METHODS OF USE THEREOF
; FILE REFERENCE: 09010-108001
; CURRENT APPLICATION NUMBER: US/10/081,872
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: US 60/270,495
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: US 60/270,496
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: US 60/291,122
; PRIOR FILING DATE: 2001-05-14
; NUMBER OF SEQ ID NOS: 321
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 153
; LENGTH: 1620
; TYPE: DNA
; ORGANISM: Environmental
US-10-081-872-153
```

## Alignment Scores:

```
Pred. No.: 3,11e-30 Length: 1620
Score: 331.00 Matches: 134
Percent Similarity: 36.63% Conservative: 77
Best Local Similarity: 23.26% Mismatches: 237
Query Match: 9.71% Indels: 129
DB: 14 Gaps: 20

US-09-843-007A-2 (1-636) x US-10-081-872-153 (1-1620)

QY 78 GlnSerTyrSerGlnArgAsnSerSerLeuLysAspIleAlaArgGluAsn 97
Db 110 CAATCATACAGCAGAAACATTAATTAAGGTCATAAAG-GCGCAACAGGAAT 168
QY 98 ProAspTrpIleLeuSerAsnLysGlnValGlyGlyValCysTyrValAspLeuPhe 116
Db 169 AAAGACCGAATTTTATGAA-----CTGTATGTAAATCTTTTAT 210
QY 117 -----AlaGlyAspLeuLysGlyLeuLysAspIleProTyrPhe 130
Db 211 GATACTGATAGCAATGACATGCTGATTTAAAGGGGTCACAAAGAACTTGATTTA 270
QY 131 -----GlnGluLeuGlyLeuThrTyrLeuHisLeuMetPro 142
Db 271 AATGATGGAAATCCAGAACAAATAATGATCTTCAATTAACCGTATCTCGATGTCCT 330
QY 143 LeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSerTyrArgAsp 162
Db 331 ATTAACACACCTCTCT-----AGTTATCAAAATATGATGTAACAGATTACTATAAT 381
QY 163 ValAsnProAlaLeuGlyThrIleGlyLeuArgGluValIleAlaLeuHisGlu 182
Db 382 ATCGATCCTCAGTATGGAAGTTTACAGATTTCCGTGAACCTAACACAGAACGCCATAAA 441
QY 183 AlaGlyIleSerAlaValValAspPheIlePheAsnHisThrSerAsnGluHisGluTyr 202
Db 442 CGCAACGTAAGGTAGTAGTAAATAGATCTTGTTATTAATCATACAGCAGTACCTCTGG 501
QY 203 AlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIlePheProAsp 222
Db 502 TTTGTGATGCATTAATAAATAAAGACAGTAAGTATCGAGATTACTATATTTGGGTGAT 561
QY 223 ArgArgMetProAspGlnTyrAspAlgThrLeuArgGluIlePheProAspGlnHisPro 242
Db 562 AAAATACAGAC-----TTAATGAAAGGCCCA 591
QY 243 GlyGly-----PheSerGlnLeuGluAspGlyArgTrpValTrpThrPheAsn 259
Db 592 TGGGGTCAACAGTATGGCACAAGCGTCGACGAGAGTATTTCTACGCAACGTTCTGG 651
QY 260 SerPheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheArgAlaMetAlaGlyGlu 279
Db 652 GAAGGATGCCGACTTAACCTATATGACACCCCTAAAGTAGAGAGAAATGATTAATATC 711
QY 280 MetLeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIle 299
Db 712 GGGAAATTTGGCTCAACAGGAGCTGATGGCTTTGCTAGATGCGCCATGCAATC 771
QY 300 TrpLysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAla 319
Db 772 TTTAAA-----GGCAAAACCTCTGAAGGAGCAAGAAATAATTAATGCTGCAATGAA 825
QY 320 PheAsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleVal 339
Db 826 TTCGCGACGCGATGAGAGAAACGAATCCAAATACGTATCTAGTTGTTGTAATATGGAT 885
QY 340 HisProAspGlnValValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnPro 359
Db 886 CAACCAAGATGATGCTCCGTAT----- 909
QY 360 LeuGlnMetAlaLeuLeuTrpAsnThrLeuAlaThrArgGluValAsnLeuHisGln 379
Db 910 -----TATCAA 915
```

Qy 380 AlaLeuThrTyArgHisAsnLeuProGluHisThrAlaTrpValAsnTy-ValArgSer 399  
Db 916 TCGTTAGATTCTACATTAACTTCGACTTAGCATATAAAATCGTTTAAATCGGTTAAAAAT 975  
Qy 400 HisAspAspIleGlyTrpThrPheAlaAspGluAspAlaAlaTyLeuGlyIleSerGly 419  
Db 976 GGTACTGATCAAGG-----GTAGCCGCGCGAGCT 1005  
Qy 420 TyrAspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAla 439  
Db 1006 GTTGCAACGGATGATTATATAAACAATATAATCAATATAAATGATGAACGTTTATA 1065  
Qy 440 ArgGly-----ValProPheGlnTyAsnProSerThrGlyAspCysArg 454  
Db 1066 ACGAATCATGACCAAAATCGTGTAAATGAGTGAATTAAT-----GGTGAT----- 1110  
Qy 455 ValSerGlyThrAlaAlaLeuValGlyLeuAlaGlnAspAspProHisAlaValAsp 474  
Db 1111 -----GTAAC 1116  
Qy 475 ArgIleLeuLeuLeuTySerIleAlaLeuSerThrGlyLeuProLeuIleTyLeu 494  
Db 1117 AAAGCAAAATCAGCAGCCTCTATTCTGTGACACTCCCTCGTGAATCCGTTCAATTATAT 1176  
Qy 495 GlyAspGluValGlyThrLeuAsnAsp-----AspAsp----- 505  
Db 1177 GCGAAGAAATCGCATGACAGCCCAAAACAGATGAGTGAATCGTGAGCCCTTCGCT 1236  
Qy 506 TrpSerGlnAspSerAsnLysSerAspSerArgTrpAlaHisArgProArgTyTrAsn 525  
Db 1237 TGGTATGAAGAT---GATAAAGAGGTCAAAGAGCTGG---GAGACTCCAGTATATAAC 1290  
Qy 526 -----GluAlaLeuTyAlaGlnArgAsnAspProSerThrAlaAlaGlyGlnIle 542  
Db 1291 ATTGATCATATAGTGTTTCAGTTGAAGCACAAGATATAACAAAAGCTTCT-----CTT 1344  
Qy 543 TyrGlnGlyLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAspGlyGly 562  
Db 1345 CTAAGCCATTATCGTAAATGATTCGTGTCTGACAGCAGCAGTGAACCTGTCAAAGGT 1404  
Qy 563 ArgLeuValThrPheAsnThrAsnAsnLysHisIleIleGlyTyIleArg----- 579  
Db 1405 AATTAGAACCTATTCTGTCAATTAATTCACAGGTGTGTGCTATATAATCGTACGATATAA 1464  
Qy 580 AsnAsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyTrpGlnThrValThrAlaHis 599  
Db 1465 AATAATCAATTCAGTGTACCATATATTTTCAGACAGCCGTTACATTAATCTTTCA 1524  
Qy 600 ThrLeuGlnAlaMetProPheLysAlaHisAspLeuIleGlyGlyLys 615  
Db 1525 AACAAAGGAAACTGATTTTCTAGTGA-----TTAGGAGCAAAA 1566

RESULT 10

US-09-815-242-8494  
; Sequence 8494, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; CURRENT APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8494  
; LENGTH: 1650  
; TYPE: DNA  
; ORGANISM: Staphylococcus aureus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1650)  
US-09-815-242-8494  
Alignment Scores:  
Pred. No.: 6,54e-30 Length: 1650  
Score: 328.50 Matches: 133  
Percent Similarity: 37.38% Conservative: 95  
Best Local Similarity: 21.80% Mismatches: 201  
Query Match: 9.64% Indels: 181  
DB: 9 Gaps: 27  
US-09-843-007a-2 (1-636) x US-09-815-242-8494 (1-1650)  
Qy 118 GlyAspLeuLysGlyLeuLysAspLysIleProTyPheGlnGluLeuGlyLeuThrTy 137  
Db 82 GGGGATTTCCTGAAGATGATTGAATAATCGACTACTTAAAGATTTAGTATCGATGTC 141  
Qy 138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyAlaVal 157  
Db 142 ATTTGGCTCAGTCCAAATGTTTAAATCACCT-----AATCATGATAATGTTATGATAT 195  
Qy 158 SerSerTyArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIle 177  
Db 196 AGTGACTACCAAGAGATTTATGGATGAATTTGGAAACGATGGAGACATTTGATCGTTTATA 255  
Qy 178 AlaAlaLeuHisGluAlaGlyIleSerAlaValAspPheIlePheAsnHisThrSer 197  
Db 256 AAGGTCTTCATGATAGAGGTATGAAGCTTATTTAGATTTAGTTGTAATCATACATCT 315  
Qy 198 AsnGluHisGluTrp-----AlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsn 215  
Db 316 GATGAACATCTCTGTTTATAGAAATCCAAATCTAGTAAAGACAAATCCCAACCGTAT--- 372  
Qy 216 PheTyTrpIlePheProAspArgArgMetProAspGlnTyAspArgThrLeuArgGlu 235  
Db 373 ---TGGTACATTTGGCAAGATCCAAAG---CCAGAT----- 402  
Qy 236 IlePheProAspGlnHisProGlyGlyPheSerGlnLeuGluAspGly----- 251  
Db 403 -----GGCTCTGAACCTCAACACTGGGAAAGTATATTTAATGATCTACATGGGAA 453  
Qy 252 -----ArgTrpValTrpThrPheAsnSerPheGlnTrpAspLeu 265  
Db 454 TATGATCTTAATCTAGCAATATATTTCCATTTATTCAGTAAACCAACACCTGATTG 513  
Qy 266 AsnTySerAsnProTrpValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsn 285  
Db 514 AATTGGGGTAATCCGGAAGTTAGAGATGCTGTATTCCGAATGATGAACCTGCTGTTGAT 573  
Qy 286 LeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTrpLysGlnMetGlyThr 305  
Db 574 AAAGCCATTGATGATTTAGATAGATGCAATTTACCATATTTAGAGAGACGTTTGAACGG 633  
Qy 306 SerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPheAsnAlaValMetArg 325

```
Db 634 GGTGACTTACCTGTACCTGAGGGTAAACATATGCCCCAGCATTTGTAGATGATGAAT 693
Qy 326 -----
Db 694 CAGCCAGGTATACAACTGCTTCAAGAGATGAAGATCGCTCATTAAGTAAGTATGAC 753
Qy 331 ValPhePheLysSerGluAlaile---ValHisProAspGlnValValGlnTyrIleGly 349
Db 754 ATTATGACTGTGTGTAAGCAATGGTGTAGCCCTGATGCTGATGCTGCTGGTGGG 813
Qy 350 GlnAspGluCysGlnIleGlyTyrAsnProLeu-----GlnMetAlaLeuLeuTyr 366
Db 814 GAAGAAATGTGTA---TTTAATATGATATTTCCAAATTTGAACATTTGGGAGCTGG 867
Qy 367 AsnThrLeuAlaThrArg---GluValAsnLeuHisGlnAlaLeuThrTyrArgHis 385
Db 868 AATAGTGGTGATCTTCATTTGTGTAAATTCGTATATAATCTGATTAATAGATGCAA 927
Qy 386 AsnLeuProGluHisThrAlaTyr---ValAsnTyrValArgSerHisAspAspIleGly 404
Db 928 AAACAACCTTGAATAATAAGGTGGAAATCGCTGTGTTATTGAATCATGACCAACCA 987
Qy 405 -----TrpThrPheAlaAspGluAspAlaAlaTyrLeuGlyIleSerGlyTyrAspHis 422
Db 988 CGTGTATCGACGTGGGTGACGATGACAAAGTATTGG-----TATGAA--- 1029
Qy 423 ArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAlaArgGlyVal 442
Db 1029 ----- 1029
Qy 443 ProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAlaAlaLeu 462
Db 1030 -----TCAGCAACAAGTCATGCACT 1050
Qy 463 ValGlyLeuAlaGlnAspAspProHisAlaValAspArgIleLysLeuLeuTyrSerIle 482
Db 1051 GTTTATTCTTCGCAACA----- 1068
Qy 483 AlaLeuSerThrGlyLeuProLeuIleTyrLeuGlyAspGluValGly----- 499
Db 1069 -----GGTACGCCATTCATTTATCAAGGTCAAGAAATTTGGTATGACGAAT 1113
Qy 500 -----ThrLeuAsnAsp-----AspAsnTyrSer--- 507
Db 1114 TATCCATTGAAAGTATTGAACATTTAAGCATGTTGCTGTTAAATATGACTATCAAAAT 1173
Qy 508 -----GlnAspSerAsnIys 512
Db 1174 GTGAAGCTCAAGGTGAGATGTAGACGCTTTACTTGGCAATATATAAGATGACACCGA 1233
Qy 513 SerAspAspSerArg-----TrpAlaHis----- 520
Db 1234 --GATAATTCTCGCACACCAATGCAATGGATGATGATGATGATGATGATGATGATGAT 1290
Qy 521 -----ArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsn 533
Db 1291 GGTGACCGTGTCCCGAGTGAATCCGAATTTAAACTATCAATGTTGCATCAATTA 1350
Qy 534 AspProSerThrAlaAlaGlyGlnIleTyrGlnGlyLeuArgHisMetIleAlaValArg 553
Db 1351 GAAGATGAGCAATTCAGTATTACAATTTTATAAGATTTA-----ATTCAATTAGA 1401
Qy 554 GlnSerAsnProArgPheAspGlyGlyArgLeuValThrPheAsnThrAsnAsnLysHis 573
Db 1402 AAATCTATGATGATATGCTATATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 1461
Qy 574 IleIleGlyTyrIleArg-----AsnAsnAlaLeuLeuAlaPheGlyAsnPheSer 590
Db 1462 GTTTTTCATCATGACGAAACATTAATGAAAGCAAGTCTTATAGTAGGTAAATCTTACT 1521
Qy 591 GluTyrProGlnThrValThrAlaHisThrLeuGlnAlaMetProPheLysAlaHisAsp 610
```

```
Db 1522 AACCCAGCAAGCAGCAATTAAC-----GTACCATT-----GAT 1554
Qy 611 LeuIleGlyGlyLysThrValSerLeuAsnGlnAspLeu-----ThrLeu 625
Db 1555 GTAAGCATGGAAGAGTGAAGCTATTATTAATGTCGCAAGCTAATTTAAACAGTTG 1614
Qy 626 GlnProTyrGlnValMetTrpLeuGluIle 635
Db 1615 CGTCCATATGACGATGTGTTATCGAATTA 1644

RESULT 11
US-09-815-242-4455
; Sequence 4455, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCES: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 4455
; LENGTH: 1638
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4455

Alignment Scores:
Pred. No.: 2,69e-29 Length: 1638
Score: 325.50 Matches: 132
Percent Similarity: 37.25% Conservative: 93
Best Local Similarity: 21.85% Mismatches: 198
Query Match: 9.49% Indels: 181
DB: Gaps: 27

US-09-843-007A-2 (1-636) x US-09-815-242-4455 (1-1638)
Qy 118 GlyAspLeuLysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyr 137
Db 82 GGGGATTACCTCGAATGATTTGATTAATTAAGACTTAAAGATTTAGGTATCGATGTC 141
Qy 138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaVal 157
Db 142 ATTTGGCTGCTCCATGTTTAAATCACCT-----AATCATGATTAATGTTATGATATT 195
Qy 158 SerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIle 177
Db 196 AGTGACTTACCAGAGATTATGATGATTAATTTGCAACGATGGAAGACTTTGATGCTTTATTA 255
Qy 178 AlaAlaLeuHisGluAlaGlyIleSerAlaValAspPheIlePheAsnHisThrSer 197
```







```
Db 1528 CCGCTCTTTGGACCGGAGCTATCCGACATCGATCCGAGTAATCCGATGCTCTATGCC 1587
Qy 577 TyrIleArgAsn-----AsnAlaLeuLeuAlaPheGlyAsnPheSerGluTyrPro 593
Db 1588 TATACCGCGACCGCAGGATCGCGAGCTATCTGGTGTAGTCACTTCAAGCGAGGCCA 1647
Qy 594 GlnThrValThr-----AlaHisThrLeuGlnAlaMetProPhe 606
Db 1648 AGAGATTTTCCAGCTTCCGACGCGATGATATTCGCGAAACCTGATTGAGAGCAGTTGG 1707
Qy 607 LysAlaHisAspLeuIleGlyGlyLysThrValSerLeuAsnGlnAspLeuThrLeuGln 626
Db 1708 CCAGCAGCTCCGCGCGCGGGCTGCAAGC-----CTTGAGCTGCAG 1749
Qy 627 ProTyrGln 629
Db 1750 CCTTGGCAG 1758

RESULT 13
US-10-081-872-191
; Sequence 191, Application US/10081872
; Publication No. US20030125534A1
; GENERAL INFORMATION:
; APPLICANT: Callen, Walter
; APPLICANT: Richardson, Toby
; APPLICANT: Frey, Gerhard
; APPLICANT: Short, Jay M.
; APPLICANT: Mathur, Eric J.
; APPLICANT: Gray, Kevin A.
; APPLICANT: Kerovio, Janne S.
; APPLICANT: Slupska, Malgorzata
; TITLE OF INVENTION: ENZYMES HAVING ALPHA AMYLASE ACTIVITY
; FILE OF INVENTION: AND METHODS OF USE THEREOF
; FILE REFERENCE: 09010-108001
; CURRENT APPLICATION NUMBER: US/10/081,872
; CURRENT FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: US 60/270,495
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: US 60/270,496
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: US 60/291,122
; PRIOR FILING DATE: 2001-05-14
; NUMBER OF SEQ ID NOS: 321
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 191
; LENGTH: 1596
; TYPE: DNA
; ORGANISM: Environmental
US-10-081-872-191

Alignment Scores:
Pred. No.: 1,83e-26 Length: 1596
Score: 300.50 Matches: 127
Percent Similarity: 36.73% Conservative: 75
Best Local Similarity: 23.09% Mismatches: 221
Query Match: 8.82% Indels: 127
DB: 14 Gaps: 20

US-09-843-007a-2 (1-636) x US-10-081-872-191 (1-1596)
Qy 112 TyrValAspLeuPhe-----AlaGlyAspLeuLysGlyLeuLys 124
Db 169 TATGTAAACTTTTACGATGCAATAAAGATGGACATGGTCACTTAAAGTCTTACA 228
Qy 125 AspLysIleProTyrPhe-----GlnGluLeuGlyLeuThr 136
Db 229 CAAAGTTGCACTTATAAATGACGCAATTTCTCATACAAAAAATGATCTTCAAGTAAC 288
Qy 137 TyrLeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAla 156
Db 289 GGGATTTGATGATGCGAGTCACCCCTTCTCT-----AGCTATCATATAATATGAT 339
```

```
Qy 157 ValSerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluVal 176
Db 340 GTAAAGGACTATTATAACATGATCCGCAATTCGCAAGATTCGCAAGATTTTCGCAAGCTG 399
Qy 177 IleAlaAlaLeuHisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThr 196
Db 400 ATGAAGAAGCAGACAAACAGAGACGTAAAGCTATTATGACCTTGTGTGAATCATACG 459
Qy 197 SerAsnGluHisGluThrAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPhe 216
Db 460 AGCAGTGAACCCCTTGGTTTCAAGTCCGCTTAAAGATAAAACAGCAAGTACAGAGAT 519
Qy 217 TyrTyrIlePheProAspArgMetProAspGlnTyrAspArgThrLeuArgGluIle 236
Db 520 TACTATATTGGGCTGATAA-----AATACCGACTTGAATGAA--- 558
Qy 237 PheProAspGlnHisProGlyGlyPheSerGln-----LeuGluAspGly 251
Db 559 -----AAAGGATCTTGGGACACAAAGTATGGCATATAAGCTCCAAACCGA 603
Qy 252 ArgTrpValTyrThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsnProTyr 271
Db 604 GAGTATTTTACCGNACGTTCTGGAGAGGATGCTGACTTAATTTAGGATAACCCGTAA 663
Qy 272 ValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeu 291
Db 664 GTAAAGAAAAGAAATCAITTAACGTCGAAAGTTTGGCTAAAGCAAGCGTTGACGGGTT 723
Qy 292 ArgMetAspAlaValAlaPheIleTyrLysGlnMetGlyThrSerCysGluAsnLeuPro 311
Db 724 CGCTAGATGCTGGCTTCATATTTTAAA-----GGTCAACAGCTGAAGCGCTAAG 777
Qy 312 GlnAlaHisAlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaProAlaVal 331
Db 778 AAAAATATCTGTGTGGATGAGTTAGAGATCGGATGAAAGAAAAAGAAATCCGAATGTA 837
Qy 332 PhePheLysSerGluAlaIleValHisProAspGlnValGlnTyrIleGlyGlnAsp 351
Db 838 TATCTAAGCGGTGAAGTATGGATCAGCTGAAGTGGTAGCTCCTTAT----- 885
Qy 352 GlnCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTyrAsnThrLeuAlaThr 371
Db 885 ----- 885
Qy 372 ArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHisThr 391
Db 886 -----TATCAATCGCTTGGATCTTTTATTAAATTTTATTAGTTAGCAGGA 927
Qy 392 AlaTrpValAsnTyrValArgSerHisAspAspIleGlyTyrThrPheAlaAspGluAsp 411
Db 928 AAAATTGTCAGCTCTGTAAAGCAGGAAATGATCAAGGA----- 966
Qy 412 AlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsnArgPhePheVal 431
Db 967 -----ATCGCCACTGCGACGACGACCAACAGATGAACCTGTTCAATATCATCAATCCA 1017
Qy 432 AsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsnProSerThrGly 451
Db 1018 AACAAATTTGATGATATCTTAAACC-----AACCATGACCAAAATCGCGTC 1065
Qy 452 AspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAlaGlnAspAspProHis 471
Db 1066 ATGAGTGAAGTGAAGCGGAT----- 1086
Qy 472 AlaValAspArgIleLysLeuLeuTyrSerIleAlaLeuSerThrGlyGlyLeuProLeu 491
Db 1087 ---GTGACAAAGCAATATCAGCTGCTTCTATCTTACTTACCTTCCTCGGCAACCCGTAT 1143
Qy 492 IleTyrLeuGlyAspGluValGlyThrLeuAsnAsp-----AspAsp----- 505
Db 1144 ATTTATTACGCTGAAGAAATCGGCATGACCGCTGAAAGCGCTGATGAATTAATCGGTGAA 1203
Qy 506 -----TrpSerGlnAspSerAsnLysSerAspSerArgTrpAlaHisArgPro 522
```

Db 1204 CCGTTCGGCTGG---TACGAAGGAAACGGACTTGGCAAAACCAAGCTGG---GAAACACCT 1257  
Qy 523 ArgTyrAsn-----GluAlaLeuTyrAlaGlnArgAsnAsp 534  
Db 1258 GTATACAAATAAGCGGGAACGGTGTCTGTAGAGACAAACCAACCAAAAGGAT--- 1314  
Qy 535 ProSerThrAlaAlaGlyGlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGln 554  
Db 1315 ---TCTTGTAAATCATTAACCGTGAATGATTCGCGTGGCTCAG 1356  
Qy 555 SerAsnProArgPheAspGlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIle 574  
Db 1357 CAGCATGAAGATGTAGTAAAGGAACCTTCAATCTATTTCAGTAGACAGTAAAGAGTT 1416  
Qy 575 IleGlyTyrIleArg-----AsnAsnAlaLeuLeuAlaPheGlyAsnPheSerGlu 591  
Db 1417 GTTGCTTATAGCGGTACGTATAGGCACTCCATTTAGTGTGATCATATATTTCAAT 1476  
Qy 592 TyrProGlnThrValThrAlaHisThrLeuGlnAlaMetProPheLysAlaHisAspLeu 611  
Db 1477 CAACCGGTAAAGATATCTGTAGCAGCGAAAGGTAAATTTGATTTTGTCTAGTGAATA--- 1533  
Qy 612 IleGlyGlyLysThrValSerLeuAsnGln 621  
Db 1534 ---GGTCTAGGAAGTCAAA---AATCAG 1557

## RESULT 14

US-10-081-872-171  
; Sequence 171, Application US/10081872  
; Publication No. US20030125534A1  
; GENERAL INFORMATION:  
; APPLICANT: Callen, Walter  
; APPLICANT: Richardson, Toby  
; APPLICANT: Frey, Gerhard  
; APPLICANT: Short, Jay M.  
; APPLICANT: Mathur, Eric J.  
; APPLICANT: Gray, Kevin A.  
; APPLICANT: Kerovuo, Janne S.  
; APPLICANT: Slupski, Malgorzata  
; TITLE OF INVENTION: ENZYMES HAVING ALPHA AMYLASE ACTIVITY  
; FILE REFERENCE: 09010-108001  
; CURRENT APPLICATION NUMBER: US/10/081,872  
; PRIOR FILING DATE: 2002-02-21  
; PRIOR APPLICATION NUMBER: US 60/270,495  
; PRIOR FILING DATE: 2001-02-21  
; PRIOR APPLICATION NUMBER: US 60/270,496  
; PRIOR FILING DATE: 2001-02-21  
; PRIOR APPLICATION NUMBER: US 60/291,122  
; PRIOR FILING DATE: 2001-05-14  
; NUMBER OF SEQ ID NOS: 321  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 171  
; LENGTH: 1431  
; TYPE: DNA  
; ORGANISM: Environmental  
US-10-081-872-171

Alignment Scores:  
Pred. No.: 1,78e-26 Length: 1431  
Score: 300.00 Matches: 122  
Percent Similarity: 36.88% Conservative: 72  
Best Local Similarity: 23.19% Mismatches: 208  
Query Match: 8.80% Indels: 124  
DB: 14 Gaps: 18

US-09-843-007a-2 (1-636) x US-10-081-872-171 (1-1431)

Qy 112 TyrValAspLeuPhe-----AlaGlyAspLeuLysGlyLeuLys 124  
Db 4 TATGTAACCTCTTTTACGATGCAATAAAGATGGACATGGTATTATAAAGGCTTACA 63

Qy 125 AspLysIleProTyrPhe-----GlnGluLeuGlyLeuThr 136  
Db 64 CAAAGCTTGGATTTATTAATGATGCAATTTCTCATCAAAAGATGATCTTCAAGTAAAC 123  
Qy 137 TyrLeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAla 156  
Db 124 GGATTTGGATGATGCGGTCAACCCCTTCCOC-----AGCTATCATATAATATGAT 174  
Qy 157 ValSerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluVal 176  
Db 175 GTAACGGACTTATTAATATGATCCGAGTATGGAATCTCGAAGATTTTCGCAAACTG 234  
Qy 177 IleAlaAlaLeuHisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThr 196  
Db 235 ATGAAGAAGACGATAAACGAGATGTAAAGTCATTATGAGACCTCGTTGTGAATCATACG 294  
Qy 197 SerAsnGluHisGlnTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPhe 216  
Db 295 AGCGTGAACACCTTGGTTTCAAGCTGCATTAAGATATAAAGATAAACAAGACGATGACAGAT 354  
Qy 217 TyrTyrIlePheProAspArgMetProAspGlnTyrAspArgThrLeuArgGluIle 236  
Db 355 TACTATATCTGGCTGATATA-----AATACCGACTTTGAATGAA--- 393  
Qy 237 PheProAspGlnHisProGlyPheSerGln-----LeuGluAspGly 251  
Db 394 -----AAGGATCTTGGGACACGCAAGTATGCGATTAAGCCCCCAACGGA 438  
Qy 252 ArgTrpValTrpThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsnProTrp 271  
Db 439 GAGTATTTTACGGAACGTTTGGGAAGGAATGCCGACTTAAATTACGATATCTCTGAA 498  
Qy 272 ValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeu 291  
Db 499 GTAAGAAAGAAATGATTAAACGTAGGAAGTTTGGCTAAAGCAAGGAGTTCGCGGTT 558  
Qy 292 ArgMetAspAlaValAlaPheIleTrpLysGlnMetGlyThrSerCysGluAsnLeuPro 311  
Db 559 CGTCTAGATGCTCGCTTCATATTTTAA-----GGCCAAACACCTGAAGCGCTAAG 612  
Qy 312 GlnAlaHisAlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaProAlaVal 331  
Db 613 AAAAAATCTCTGTGTGGGAATGAAATTTAGAGATGCAATGAAAGGAAACCCCTAACGTA 672  
Qy 332 PhePheLysSerGluAlaIleValHisProAspGlnValGlnTyrIleGlyGlnAsp 351  
Db 673 TATCTAACGGGTGAAGTATGGATGCAACCGAGTAGTAGCTCTTAC----- 720  
Qy 352 GluCysGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTrpAsnThrLeuAlaThr 371  
Db 720 ----- 720  
Qy 372 ArgGluValAsnLeuLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHisThr 391  
Db 721 -----TATCATTCGCTGATCTTTTATTAACTTTGATTAGCAGGA 762  
Qy 392 AlaTrpValAsnTyrValArgSerHisAspAspIleGlyTrpThrPheAlaAspGluAsp 411  
Db 763 AAGATTGTAAACTCTGTAAATCAAGGAATGATCAAGGA----- 801  
Qy 412 AlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsnArgPhePheVal 431  
Db 802 -----ATCGGACTCGACAGCGCAACGAGTATGTTTCAAAATCATACATCAATCCA 852  
Qy 432 AsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsnProSerThrGly 451  
Db 853 AATAAAATTGACGGTATTTTCTTAACC-----AACCATGACCAAAATCGCGTC 900  
Qy 452 AspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAlaGlnAspAspProHis 471  
Db 901 ATGAGTGAGCTAAGCGCGCAT----- 921  
Qy 472 AlaValAspArgIleLysLeuLeuTyrSerIleAlaLeuSerThrGlyGlyLeuProLeu 491

```

Db 922 ---GTGAATAAGCAAGTCAGCTGCTCTACTTCTTACTTACGCTTCTCGCAACCGGTAT 978
Qy 492 lletyLeuGlyAspGluValGlyThrLeuAsnAsp-----AspAsp----- 505
Db 979 AATTATTACGGTGAAGAAATGCGCAGACCGGTGAAACCGCTGATGATTAATCCGTGAA 1038
Qy 506 -----TTPSerGlnAspSerAsnLysSerAspAspSerArgTTPAlaHisArgPro 522
Db 1039 CGGTTCCGCTGG---TACGAAGMAACGGACTTGGACAAACCGCTGG---GAAACACT 1092
Qy 523 ArgTyrAsn-----GluAlaLeuTyrAlaGlnArgAsnAsp 534
Db 1093 GTATACAAAGGCGGCAACCGGTGCTGTAGAGCAACAAACAAAGGAC--- 1149
Qy 535 ProSerThrAlaAlaGlyGlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGln 554
Db 1150 -----TCTTTGTTAATCATTAACCGTGAATGATTTCGGCGGTCTAG 1191
Qy 555 SerAsnProArgPheAspGlyArgLeuValThrPheAsnThrAsnAsnLysHisIle 574
Db 1192 CAGCAGAGAGATTAGTAAGCAACCGCTTCAATCTATTTCATGACAGACAGTAAGAGATC 1251
Qy 575 IleGlyTyrIleArg-----AsnAsnAlaLeuAlaPheGlyAsnPheSerGlu 591
Db 1252 GTTGCCTATAGCGGCACGTATTAAGCAATCGATTAGCGTATCATTAATATTTCRAAT 1311
Qy 592 TyrProGlnThrValThr 597
Db 1312 CAACCGTAAAGATATCT 1329

RESULT 15
US-10-081-872-167
; Sequence 167, Application US/10081872
; Publication No. US20030125534A1
; GENERAL INFORMATION:
; APPLICANT: Richardson, Walter
; APPLICANT: Callen, Toby
; APPLICANT: Prey, Gerhard
; APPLICANT: Short, Jay M.
; APPLICANT: Mathur, Eric J.
; APPLICANT: Gray, Kevin A.
; APPLICANT: Kerovuo, Janne S.
; APPLICANT: Slupska, Malgorzata
; TITLE OF INVENTION: ENZYMES HAVING ALPHA AMYLASE ACTIVITY
; TITLE OF INVENTION: AND METHODS OF USE THEREOF
; FILE REFERENCES: 09010-108001
; CURRENT APPLICATION NUMBER: US/10/081,872
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: US 60/270,495
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: US 60/270,496
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: US 60/291,122
; PRIOR FILING DATE: 2001-05-14
; NUMBER OF SEQ ID NOS: 321
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 167
; LENGTH: 1596
; TYPE: DNA
; ORGANISM: Environmental
US-10-081-872-167

Alignment Scores:
Pred. No.: 2,11e-26 Length: 1596
Score: 300.00 Matches: 122
Percent Similarity: 36.88% Conservative: 72
Best Local Similarity: 23.19% Mismatches: 208
Query Match: 8.80% Indels: 124
DB: 14 Gaps: 18

US-09-843-007A-2 (1-636) x US-10-081-872-167 (1-1596)
```

```

Qy 112 TyrValAspLeuPhe-----AlaGlyAspLeuLysGlyLeuLys 124
Db 169 TATGTAAACTCTTTTACGATGCAATAAAGATGACATGCTGATTTAAAGGCTTTACA 228
Qy 125 AsplysIleProTyrPhe-----GlnGluLeuGlyLeuThr 136
Db 229 CAAAGTTGGATTATTTAAATGATGGCAATTCTCATCAAGAAGATGATCTTCAAGTAAAC 288
Qy 137 TyrLeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAla 156
Db 289 GGGATTGGATGATGCGGTCAACCTTCTCCC-----AGCTATCATATAATATCAT 339
Qy 157 ValSerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluVal 176
Db 340 GTAACGGACTATTATAATATTGATCCCGCATGCAAAATCTCGAAGATTTTCGCAACTG 399
Qy 177 IleAlaAlaLeuHisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThr 196
Db 400 ATGAAGNAGCAGATAAACGAGATGTAAGTCTATTATGGACCTGTTGTGATCATACG 459
Qy 197 SerAsnGluHisGluTTPAlaGlnArgCysAlaAlaGlyAspProLeuPheAsnAsnPhe 216
Db 460 AGCAGTGAACACCTTGGTTTCAAGCTGCATTAAGATAAAACAGCAAGTACAGAGAT 519
Qy 217 TyrTyrIlePheProAspArgMetProAspGlnTyrAspArgThrLeuArgGluIle 236
Db 520 TACTATATCTGGCTGATPAA-----AATCCGACTTGAATGAA--- 558
Qy 237 PheProAspGlnHisProGlyGlyPheSerGln-----LeuGluAspGly 251
Db 559 -----AAAGGATCTTGGGACAGCAAGATATGGCATATAAGCCCAACCGGA 603
Qy 252 ArgTTPValTTPThrPheAsnSerPheGlnTTPAspLeuAsnTyrSerAsnProTTP 271
Db 604 GAGTATTTTACGGAACCTTTTGGGAGGAATGCGGACTTTAAATTACGATATCTCTGAA 663
Qy 272 ValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAspIleLeu 291
Db 664 GTAAGAAAGAAATGATTAACTAGGAAAGTTTGGCTAAAGCAAGAGTTGACGGGTTC 723
Qy 292 ArgMetAspAlaValAlaPheIleTTPlysGlnMetGlyThrSerCysGluAsnLeuPro 311
Db 724 CGCTAGATGCTGGCTTCATATTTTAA-----GCCCAACACCTGAAGCGGCTAAG 777
Qy 312 GlnAlaHisAlaLeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaProAlaVal 331
Db 778 AAAAATCTCTCTGCTGGGCAATGATTTAGAGATGCAATGAAAGAGAAACCCCTAACGTA 837
Qy 332 PhePheLysSerGluAlaIleValHisProAspGlnValGlnTyrIleGlyGlnAsp 351
Db 838 TATCTAACGGGTGAAGTAGGATCAACCGAAGTAGTGTCTCTTAC----- 885
Qy 352 GluCysGlnIleGlyTyrAsnProLeuGlnMetalLeuLeuTTPAsnThrLeuAlaThr 371
Db 885 ----- 885
Qy 372 ArgGluValAsnLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHisThr 391
Db 886 -----TATCAATCGCTTGTATTTTATTTAACTTTTATTTAGCAGGA 927
Qy 392 AlaTTPValAsnTyrValArgSerHisAspAspIleGlyTTPThrPheAlaAspGluAsp 411
Db 928 AAGATTGTAACCTCTGTAATTCAGGAATGATCAGGA----- 966
Qy 412 AlaAlaTyrLeuGlyIleSerGlyTyrAspHisArgGlnPheLeuAsnArgPhePheVal 431
Db 967 -----ATCGCGACTGCGACGCGCAACGGATGAACCTTCAAAATCATACATCA 1017
Qy 432 AsnArgPheAspGlySerPheAlaArgGlyValProPheGlnTyrAsnProSerThrGly 451
Db 1018 AATAAAATGACGGTATTTTCTTAACC-----AACCATGACCAATATCGCGTC 1065
Qy 452 AspCysArgValSerGlyThrAlaAlaLeuValGlyLeuAlaGlnAspAspProHis 471
```

```
Db      1066  ATGAGTGAAGCTAAGCGCGAT----- 1086
Qy      472  AlaValAspArgIleLeuLeuTyrSerIleAlaLeuSerThrGlyGlyLeuProLeu 491
Db      1087  ---GTGATAAAGCAAGTCACTGCCTCTATCTTACTTACGCTTCTGGCAACCGTAT 1143
Qy      492  IleTyrLeuGlyAspGluValGlyThrLeuAsnAsp-----AspAsp----- 505
Db      1144  ATTTATTACGGTGAAGAAATCGGCATACCGGTGAAAAGCCGTGATGATTAATCCGTGAA 1203
Qy      506  -----TTPSerGlnAspSerAsnLysSerAspAspSerArgTTPAlaHisArgPro 522
Db      1204  CCGTTCGCTGG---TAGGAAGGAACGGACTTGGCAAAACCACTGG---GAAACACCT 1257
Qy      523  ArgTyrAsn-----GluAlaLeuTyrAlaGlnArgAsnAsp 534
Db      1258  GTATACAAAGCGGCAACGGCGTGTCTGTAGAGCACAAACAAACAAAGGAC--- 1314
Qy      535  ProSerThrAlaAlaGlyGlnIleTyrGlnGlyLeuArgHisMetIleAlaValArgGln 554
Db      1315  -----TCTTTGTAAATCATTCACCGTGAATGATTCGCGTCGCTCAG 1356
Qy      555  SerAsnProArgPheAspGlyGlyArgLeuValThrPheAsnThrAsnAsnLysHisIle 574
Db      1357  CAGCACGAAGAGTTAGTAAGGAACGCTTCAATCTATTTCAGTAGACAGTAAGAGTC 1416
Qy      575  IleGlyTyrIleArg-----AsnAsnAlaLeuLeuAlaPheGlyAsnPheSerGlu 591
Db      1417  GTTCCTATAGCCCGCACGTATAAAGGCAATCGATTAGCGTGTATCATATATTCAAT 1476
Qy      592  TyrProGlnThrValThr 597
Db      1477  CAACCGGTAAAGTATCT 1494
```

Search completed: November 9, 2003, 01:50:34  
Job time : 6659 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.  
OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: November 8, 2003, 19:48:18 ; Search time 2885 Seconds  
(without alignments)  
5357.937 Million cell updates/sec

Title: US-09-843-007A-2  
Perfect score: 3408  
Sequence: 1 MLPTQGVLIQLVLTTRIL.....VSLNQDLTLQPYQVMWLEIA 636  
Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Zgapop 6.0, Zgapext 7.0  
Delop 6.0, Delext 7.0  
Searched: 22781392 seqs, 12152238056 residues  
Total number of hits satisfying chosen parameters: 45562784  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2 1/USPTO spool p/US09843007/runat 07112003 170420 25360/app query.fasta 1.775  
-DB=EST -OPMT=fastap -SUFFIX=p2n.rest -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THL\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09843007/CGCN 1.1 3549 @runat 07112003 170420 25360 -NCPU=6 -ICPU=3  
-NO MVAP -LARGSEQQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_kod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gesi:\*

29: gb\_gse2:\*  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	285.5	8.4	659	9	AI107032	AI107032 GH06318.5
2	278.5	8.2	641	9	AI292849	AI292849 GH15856.5
3	277	8.1	679	12	BM56725	BM56725 170006873
4	276.5	8.1	626	9	AI297229	AI297229 LP11488.5
5	274.5	8.1	638	9	AI402332	AI402332 GH21508.5
6	272.5	8.0	1981	28	BH770895	BH770895 LLMGT-ag62
7	271.5	8.0	673	12	BM610365	BM610365 170006871
8	271.5	8.0	712	12	BM646129	BM646129 170006873
9	265	7.8	785	29	BZ578463	BZ578463 msh2_5885
10	263	7.7	587	12	BM619567	BM619567 170006874
11	263	7.7	711	12	BM652118	BM652118 170006873
12	262.5	7.7	704	12	BM628529	BM628529 170006874
13	262.5	7.7	706	12	BM650445	BM650445 170006873
14	262.5	7.7	912	11	CNS08PQT	EX022785 Single re
15	260.5	7.6	694	12	BM579790	BM579790 170006872
16	257	7.5	611	12	BI507533	BI507533 BB17000687A
17	256.5	7.5	648	12	BI583466	BI583466 RH22083.5
18	256.5	7.5	703	12	BM587500	BM587500 170006873
19	254.5	7.5	2271	11	AK052623	AK052623 Mus muscu
20	252.5	7.4	884	28	BH149112	BH149112 BNTQ104TF
21	252	7.4	718	12	BM655933	BM655933 170006873
22	251.5	7.4	684	12	BM635945	BM635945 170006871
23	251.5	7.4	698	12	BM616688	BM616688 170006873
24	251.5	7.4	711	12	BM641301	BM641301 170006873
25	250.5	7.4	604	9	AI238307	AI238307 GH14330.5
26	250.5	7.4	684	12	BM620803	BM620803 170006874
27	250	7.3	650	12	BJ329346	BJ329346 BJ329346
28	249.5	7.3	684	12	BM634536	BM634536 170006875
29	248.5	7.3	615	12	BI587736	BI587736 RH2884.5
30	248	7.3	688	12	BM624147	BM624147 170006874
31	248	7.3	693	12	BM637841	BM637841 170006875
32	248	7.3	693	12	BM604234	BM604234 170006870
33	248	7.3	693	12	BM622334	BM622334 170006874
34	248	7.3	694	12	BM597646	BM597646 170006875
35	246.5	7.2	621	12	BI634521	BI634521 RH3895.5
36	246.5	7.2	628	12	BJ365771	BJ365771 BJB365771
37	246	7.2	665	12	BM652834	BM652834 170006873
38	245.5	7.2	616	12	BI564072	BI564072 RH35755.5
39	245.5	7.2	623	9	AI389773	AI389773 GH21031.5
40	245.5	7.2	681	12	BM614246	BM614246 170006871
41	244	7.2	696	12	BM613854	BM613854 170006871
42	244	7.2	696	12	BM652741	BM652741 170006873
43	243	7.1	696	12	BM651858	BM651858 170006873
44	243	7.1	699	12	BM625632	BM625632 170006874
45	242.5	7.1	590	12	BJ324954	BJ324954 BJB324954

ALIGNMENTS

RESULT 1  
AI107032  
LOCUS GH06318.5prime GH Drosophila melanogaster head p0T2 Drosophila  
DEFINITION melanogaster cDNA clone GH06318 5prime similar to V00204: Lvph  
P890002570 PID:g7777 SWISS-PROT:P07190, mRNA sequence.  
659 bp mRNA linear EST 19-APR-2001  
ACCESSION AI107032  
VERSION AI107032.1 GI:3475967  
KEYWORDS EST.  
SOURCE Drosophila melanogaster (fruit fly)  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 659)  
Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,  
Lewis, S., and Rubin, G.M.  
BDGP/HEMI Drosophila EST Project  
Unpublished  
Contact: Stapleton, M.  
BDGP  
Lawrence Berkeley National Lab  
One Cyclotron Rd., Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: <http://www.fruitfly.org/EST>; [est@fruitfly.berkeley.edu](mailto:est@fruitfly.berkeley.edu)  
hit genomic sequence DS04362; hit genomic sequence DS06189  
Plate: 63 row: B column: 6  
High quality sequence stop: 515.  
Location/Qualifiers  
1. .659  
/organism="Drosophila melanogaster"  
/mol\_type="mRNA"  
/db\_xref="taxon:7227"  
/clone="GH06318"  
/sex="male and female"  
/dev\_stage="adult"  
/clone\_lib="GH Drosophila melanogaster head pOT2"  
/note="Organ: head; Vector: pOT2; Site 1: EcoRI; Site 2:  
XhoI; Sized fractionated cDNAs were directly ligated into  
pOT2. Plasmid cDNA library."  
XhoI; Sized fractionated cDNAs were directly ligated into  
pOT2. Plasmid cDNA library."

BASE COUNT 148 a 160 c 191 g 159 t 1 others  
ORIGIN  
Alignment Scores:  
Pred. No.: 6,16e-23 Length: 659  
Score: 285.50 Matches: 65  
Percent Similarity: 54.75% Conservative: 33  
Best Local Similarity: 36.31% Mismatches: 76  
Query Match: 8.38% Indels: 5  
DB: 9 Gaps: 3

US-09-843-007A-2 (1-636) x AJ107032 (1-659)  
Qy 118 GlyAspLeuHisGluAlaValAspLeuGlyThrLeuTyr 137  
Db 129 GCGGATTTGAAGGAAATCCCTCCAGCTGCCTATCTGGCGGACACTGGGATCAGGCC 188

Qy 138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAlaVal 157  
Db 189 ACCTGGTTGAGTCCCATTTTTCAGTCGCC-----ATGATTGACTTTGGCTATGATATA 242

Qy 158 SerSerTyrArgAspValAsnProAlaLeuGlyThrLeuGlyAspLeuArgGluValle 177  
Db 243 TCGGATTACAGGCAATCCAGCGGAGTATGGCACCATGCGAGGATTCGAGGAGCTGATC 302

Qy 178 AlaAlaLeuHisGluAlaValSerAlaValAspPheLeuHisThrSer 197  
Db 303 GACACGGCTTCGAGCTGGGATCAGGTGTCTGGACTTTGTCCGGAATCAGCTCG 362

Qy 198 AsnGluHisGluThrAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyr 217  
Db 363 GATCAGCATGAGTGTTCAGAGTCTGCGCAAGGAGCGGGTACGAGGATTC--- 419

Qy 218 TyrIlePheProAspArgMetProAspGlnTyrAspArgThrLeuArgGluValle 237  
Db 420 TATGTGTGCACATGGCATGTCTCAGGAGATGCGACTCGGTCCACCCACACTGG 479

Qy 238 ProAspGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArg-----TrpValTrp 255  
Db 480 CCATCGGTCTTACGGATCCGCTTGGGAGTGGCAGAGGTCTGTCAGCAGTATTACCTG 539

Qy 256 ThrThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheArgala 275  
Db 540 CACCAGTTCACCAAGAGCGCGGACTTGAACATATCGCAATCCCAAGAGTGTTCAGGCC 599

Qy 276 MetAlaGlyGluMetLeuPheLeuAlaAsnLeuGlyValAspLeuArgMetAsp 294  
Db 600 ATGGATGATGTGTGCTTTCTGCTGCACAGGCGGTANCTGGTTTTCGATCATCAT 656

RESULT 2  
AL292849 641 bp mRNA linear EST 19-APR-2001  
LOCUS  
DEFINITION  
GHI5856.Sprime GH Drosophila melanogaster head pOT2 Drosophila  
melanogaster cDNA clone GHI5856 Spprime similar to V00204: LvpH  
F8gn0002570 PID:G7777 SWISS-PROT:P07190, mRNA sequence.

ACCESSION  
AL292849  
VERSION  
AL292849.1 GI:3942256  
KEYWORDS  
EST.  
SOURCE  
Drosophila melanogaster (fruit fly)

ORGANISM  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
1 (bases 1 to 641)  
AUTHORS  
Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,  
Lewis, S., and Rubin, G.M.

TITLE  
BDGP/HEMI Drosophila EST Project  
JOURNAL  
Unpublished  
COMMENT  
Contact: Stapleton, M.  
BDGP

Lawrence Berkeley National Lab  
One Cyclotron Rd. Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: <http://www.fruitfly.org/EST>; [est@fruitfly.berkeley.edu](mailto:est@fruitfly.berkeley.edu)  
Plate: 158 row: E column: 8  
High quality sequence stop: 477.  
Location/Qualifiers  
1. .641  
/organism="Drosophila melanogaster"  
/mol\_type="mRNA"  
/db\_xref="taxon:7227"  
/clone="GHI5856"  
/sex="male and female"  
/dev\_stage="adult"  
/clone\_lib="GH Drosophila melanogaster head pOT2"  
/note="Organ: head; Vector: pOT2; Site 1: EcoRI; Site 2:  
XhoI; Sized fractionated cDNAs were directly ligated into  
pOT2. Plasmid cDNA library."

BASE COUNT 147 a 157 c 186 g 151 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 3.99e-22 Length: 641  
Score: 278.50 Matches: 63  
Percent Similarity: 54.91% Conservative: 32  
Best Local Similarity: 36.42% Mismatches: 73  
Query Match: 8.17% Indels: 5  
DB: 9 Gaps: 3

US-09-843-007A-2 (1-636) x AJ292849 (1-641)  
Qy 118 GlyAspLeuHisGluAlaValAspLeuGlyThrLeuTyr 137  
Db 130 GCGGATTTGAAGGAAATCCCTCCAGCTGCCTATCTGGCGGACACTGGGATCAGGCC 189

Qy 138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAlaVal 157  
Db 190 ACCTGGTTGAGTCCCATTTTTCAGTCGCC-----ATGATTGACTTTGGCTATGATATA 243

Qy 158 SerSerTyrArgAspValAsnProAlaLeuGlyThrLeuGlyAspLeuArgGluValle 177  
Db 244 TCGGATTACAGGCAATCCAGCGGAGTATGGCACCATGCGAGGATTCGAGGAGCTGATC 303

Qy 178 AlaAlaLeuHisGluAlaValSerAlaValAspPheLeuHisThrSer 197  
Db 304 GACACGGCTTCGAGCTGGGATCAGGTGTCTGGACTTTGTCCGGAATCAGCTCG 363



FEATURES  
source

Location/Qualifiers  
1..626  
/organism="Drosophila melanogaster"  
/mol\_type="mRNA"  
/db\_xref="taxon:7227"  
/clone="LP11488"  
/sex="male and female"  
/dev\_stage="larvae-pupae"  
/lab\_host="DHS-alpha"  
/clone\_lib="LP Drosophila melanogaster larval-early pupal  
POT2"  
/note="Organ: whole body; Vector: POT2; Site: 1: EcoRI;  
Site: 2: XhoI; Sized fractionated cDNAs were directly  
ligated into POT2. Plasmid cDNA library."

BASE COUNT  
ORIGIN  
142 a 153 c 181 g 150 t

Alignment Scores:  
Pred. No.: 6,65e-22 Length: 626  
Score: 276.50 Matches: 62  
Percent Similarity: 55.29% Conservative: 32  
Best Local Similarity: 36.47% Mismatches: 71  
Query Match: 8.11% Indels: 5  
DB: 9 Gaps: 3

US-09-843-007A-2 (1-636) x AI297229 (1-626)

QY 118 GlyAspLeuGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyr 137  
Db 124 GCGATTTGAAGGAATCACTCCAGCTGGGTATCTGGCGACACTGGCATCGAGGCC 183  
QY 138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAlaVal 157  
Db 184 ACCTGTTGAGTCCCATTTTTCAGTCGCCC-----ATGATTGACTTTTGGCTATGATATA 237  
QY 158 SerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIle 177  
Db 238 TCGATTACAGGCAATCCAGCGGAGTATGGCACCATGCAGGATTTTCAGGAGCTGATC 297  
QY 178 AlaAlaLeuHisGluAlaGlyIleSerAlaValAlaValAspPheIlePheAsnHisThrSer 197  
Db 298 GACACGGCTTTTCAGCTGGGCAATCAAGGTTGTTCTGGACTTTTGTGCGGAATCACAGCTCG 357  
QY 198 AsnGluHisGlnTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyr 217  
Db 358 GATCAGCATGAGTGGTTCAAGAGTCTGGGCAAGGAGCGCGGTTACAGAGATTTC--- 414  
QY 218 TyrIlePheProAspArgMetProAspGlnTyrAspArgThrLeuArgGluIlePhe 237  
Db 415 TATGTCTGGCAGCATGCATCGTCCAGGAGATGGCACTCGGGTGGCCACCCCAACTGG 474  
QY 238 ProAspGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArg-----TrpValTrp 255  
Db 475 CCATCGGTGTTTACGATCCGCTTGGAGTGGCAGAGGTTCTGAGCAGTATTACCTG 534  
QY 256 ThrThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheArgAla 275  
Db 535 CACCAGTTTCAACCAAGGAGGAGCGGACTTGAACATATCGCAATCCCAAGTGGTTACGGCC 594  
QY 276 MetAlaGlyGluMetLeuPheGluAlaAsn 285  
Db 595 ATGGATGATGTTGTTGCTTTCTTGTGCTCAAC 624

RESULT 5  
AI402332  
LOCUS  
DEFINITION  
GH21508.5prime GH Drosophila melanogaster head POT2 Drosophila  
melanogaster cDNA clone GH21508 5prime similar to V00204: LvpH  
F8gn0002570 PID:g7777 SWISS-PROT:P07190, mRNA sequence.  
AI402332  
AI402332.1 GI:4245419  
EST.

ACCESSION  
VERSION  
KEYWORDS

SOURCE  
ORGANISM

Drosophila melanogaster (fruit fly)  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 638)  
Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,  
Lewis,S. and Rubin,G.M.  
BDGP/HMI Drosophila EST Project  
Unpublished  
Contact: Stapleton, M.  
BDGP  
Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: <http://www.fruitfly.org/EST>, [est@fruitfly.berkeley.edu](mailto:est@fruitfly.berkeley.edu)  
hit genomic sequence AC06240  
Plate: 215 row: A column: 8  
High quality sequence stop: 551.

FEATURES  
source

Location/Qualifiers  
1..638  
/organism="Drosophila melanogaster"  
/mol\_type="mRNA"  
/db\_xref="taxon:7227"  
/clone="GH21508"  
/sex="male and female"  
/dev\_stage="adult"  
/lab\_host="DHS - alpha"  
/clone\_lib="GH Drosophila melanogaster head POT2"  
/note="Organ: head; Vector: POT2; Site: 1: EcoRI; Site: 2:  
XhoI; Sized fractionated cDNAs were directly ligated into  
POT2. Plasmid cDNA library."

BASE COUNT  
ORIGIN  
145 a 156 c 186 g 151 t

Alignment Scores:  
Pred. No.: 1.18e-21 Length: 638  
Score: 274.50 Matches: 62  
Percent Similarity: 54.55% Conservative: 32  
Best Local Similarity: 36.05% Mismatches: 73  
Query Match: 8.05% Indels: 5  
DB: 9 Gaps: 3

US-09-843-007A-2 (1-636) x AI402332 (1-638)

QY 118 GlyAspLeuLysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyr 137  
Db 130 GCGATTTGAAGGAATCACTCCAGCTGGGTATCTGGCGACACTGGCATCGAGGCC 189  
QY 138 LeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAlaVal 157  
Db 190 ACCTGTTGAGTCCCATTTTTCAGTCGCCC-----ATGATTGACTTTTGGCTATGATATA 243  
QY 158 SerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIle 177  
Db 244 TCGATTACAGGCAATCCAGCGGAGTATGGCACCATGCAGGATTTTCAGGAGCTGATC 303  
QY 178 AlaAlaLeuHisGluAlaGlyIleSerAlaValAlaValAspPheIlePheAsnHisThrSer 197  
Db 304 CACACGGCTTTTCAGCTGGGCAATCAAGGTTGTTCTGGACTTTTGTGCGGAATCACAGCTCG 363  
QY 198 AsnGluHisGlnTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyr 217  
Db 364 GATCAGCATGAGTGGTTCAAGAGTCTGGGCAAGGAGCGCGGTTACAGAGATTTC--- 420  
QY 218 TyrIlePheProAspArgMetProAspGlnTyrAspArgThrLeuArgGluIlePhe 237  
Db 421 TATGTCTGGCAGCATGCATCGTCCAGGAGATGGCACTCGGGTGGCCACCCCAACTGG 480  
QY 238 ProAspGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArg-----TrpValTrp 255  
Db 481 CCATCGGTGTTTACGATCCGCTTGGAGTGGCAGAGGTTCTGAGCAGTATTACCTG 540

RESULT 5  
AI402332  
LOCUS  
DEFINITION  
GH21508.5prime GH Drosophila melanogaster head POT2 Drosophila  
melanogaster cDNA clone GH21508 5prime similar to V00204: LvpH  
F8gn0002570 PID:g7777 SWISS-PROT:P07190, mRNA sequence.  
AI402332  
AI402332.1 GI:4245419  
EST.

ACCESSION  
VERSION  
KEYWORDS



QY 256 ThrThrPheAsnSerPheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheArgAla 275  
 DB 541 CACCAGTTTCCACGAGCAGCGGACCTGAACATATCGCAATCCCAAGTGTTCAGGCC 600  
 QY 276 MetAlaGlyGluMetLeuPheLeuAlaLeuGly 287  
 DB 601 ATGGATGATGTGTTCTGTTCTGGCTCAACAAGGC 636

RESULT 6  
 BH770895/c  
 LOCUS BH770895 1981 bp DNA linear GSS 01-MAY-2002  
 DEFINITION LLMGTAG629 MG1363 Random Sequence Tag Library Lactococcus lactis  
 subsp. cremoris genomic, genomic survey sequence.  
 ACCESSION BH770895  
 VERSION BH770895.1 GI:20373852  
 KEYWORDS GSS.  
 SOURCE Lactococcus lactis subsp. cremoris  
 ORGANISM Lactococcus lactis subsp. cremoris  
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 Lactococcus.

REFERENCE  
 1 (bases 1 to 1981)  
 Authors Bolotin, A., Ehrlich, S.D. and Sorokin, A.  
 TITLE Studies of genomes of dairy bacteria Lactococcus lactis  
 JOURNAL Sci. Aliments, (2002) In press  
 COMMENT Contact: Sorokin A  
 Genetique Microbiome  
 INRA  
 CRJ INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France  
 Tel: 33 1 34 65 25 16  
 Fax: 33 1 34 65 25 21  
 Email: sorokine@jouy.inra.fr  
 best homologue in strain IL1403 is amyY (90%)  
 Class: shotgun  
 High quality sequence start: 30  
 High quality sequence stop: 1953.  
 Location/Qualifiers  
 1..1981  
 /organism="Lactococcus lactis subsp. cremoris"  
 /mol\_type="genomic DNA"  
 /strain="MG1363"  
 /db\_xref="taxon:1359"  
 /clone\_lib="MG1363 Random Sequence Tag Library"  
 /note="Vector: pSGM2; Site 1: Smar; Library of  
 chromosomal fragments of L.lactis strain MG1363 was  
 prepared by partial AluI digestion or by sonication."  
 BASE COUNT 574 a 429 c 335 g 643 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1.22e-20 Length: 1981  
 Score: 272.50 Matches: 136  
 Percent Similarity: 38.42% Conservative: 93  
 Best Local Similarity: 22.82% Mismatches: 241  
 Query Match: 8.00% Indels: 127  
 DB: 28 Gaps: 22

US-09-843-007A-2 (1-636) x BH770895 (1-1981)

QY 75 GlnAlaTrpGlnSerTyrSerGlnArgAsnSerSerLeuLysAspIleAlaArg 94  
 DB 1885 AGGCTTCTCAAAAGTAACTTCAAAAGATTCATCAGTAAAGCGGCTGTCAGTCAA 1826  
 QY 95 GluAsnAsnProAspTrpIleLeuSerAsnLysGlnValGlyValCysTyrValAsp 114  
 DB 1825 AAGGTTGATACGAGTTTGTATCGTAATTTTATGAATA-----TTTACGAGT 1778  
 QY 115 LeuPheAla-----GlyAspLeuLysGlyLeuLysAspLysIle 127  
 DB 1777 TCATTGGCGATTCAATCATGATGAGTGGGCGGATTAAATGGGTCACCGACATTTA 1718  
 QY 128 ProTyrPheGln-----GluLeuGlyLeuThrTyrLeuHis 139  
 DB 1717 GATTATCTCAATACGGGGAATAAAATTTCTACACAGATTTGAAAGTCCAAAGGACTTTGG 1658

QY 140 LeuMetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAlaValSerSer 159  
 DB 1657 ATGACCCCAATCTTTGCTAGTCCA-----AGCTATCATGCTATGATGTCCAGCAAT 1607  
 QY 160 TyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAla 179  
 DB 1606 TACGAAGAAATTAATCTCTAAATTTGGCACAATGCTGATTTGAAATCTGATGCTCTCAG 1547  
 QY 180 LeuHisGluAlaGlyIleSerAlaValAlaAspPheIlePheAsnHisThrSerAsnGlu 199  
 DB 1546 GCTAAATAACGAGGAATTCGGTGATTTGGACATCGCGTTATCCACACGCGACTGAC 1487  
 QY 200 HisGlnTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTyrIle 219  
 DB 1486 AATCTCTGGTTTCAAAAAGCAGTACGTGCGCACAAGAAATATATATGCGTTATTAT 1433  
 QY 220 PheProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAsp 239  
 DB 1432 -----AATGTCGACACAGCAAGAG----- 1409  
 QY 240 GlnHisProGlyPheSerGlnLeuGluAspGlyArgTrpValTrpThrThrPheAsn 259  
 DB 1408 -----GGTTATAGCTTAGCCAGCAACGCAATACTATGATCAGAGTTGAT 1361  
 QY 260 SerPheGlnTrpAspLeuAsnTyrSerAsnProTrpValPheArgAlaMetAlaGlyGlu 279  
 DB 1360 AAATCGATCGCAGACCTCAATCTTCCCAATCTCGAGTCAAAAAGAAATTTGAAAAATT 1301  
 QY 280 MetLeuPheLeuAlaAsnLeuGlyValAlaIleLeuArgMetAspAlaVal-AlaPheIle 299  
 DB 1300 ACGAAATTTGGTTAGACAAAGGTGTTCTCGTTTACGTTTGGACGCTGCGTTTTTAT 1241  
 QY 299 eTrpLysGlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAl 319  
 DB 1240 TTCTCAAAATGATGATAAAAAACGACGCGCTTTACCAAAATGGCTGCAGCAG---ATACGCTC 1184  
 QY 319 aPheAsnAlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleVal 339  
 DB 1183 AAAAGCAAAATCCAAAGCTTATTAGTCGGTGAAGTTTTTTTCAA----- 1138  
 QY 339 LHisProAspGlnValValGlnTyrIleGlyGlnAspGluCysGlnIleGlyTyrAsnPr 359  
 DB 1137 -CGGCTTCGGCAATGTATGATATTTATCATGTCAG-----GTGTTGATCTTTATTTGAT 1085  
 QY 359 oLeuGlnMetAlaLeuLeuTrpAsnThrLeuAlaThrArgGluValAsnLeuHisGlu 379  
 DB 1084 TTTCCTCAATG-----CTTTGCAATCGAGTTCGAGTCCGATTAATTCGGCTA----- 1039  
 QY 379 nAlaLeuThrTyrArgHisAsnLeuProGluHisThrAlaTrpValAsnTyrValArgSe 399  
 DB 1039 ----- 1039  
 QY 399 rHisAspAspIleGlyTrpThrPheAlaAspGluAspAlaAlaTyrIleuGlyIleSerGl 419  
 DB 1038 -----TGGTCTTGGCGAAGCGCTCATTTGTTGGTCTCAG-CAAAATCGAAGC 996  
 QY 419 yTyrAspHisArgGlnPheLeuAsnArgPhePheValAsnArgPheAspGlySerPheAl 439  
 DB 995 TTGGGACCAAGAA-----ATTCATGCTCACAACCGCAAGC 960  
 QY 439 aArgGlyValProPheGlnTyrAsnProSerThrGlyAspCysArgValSerGlyThrAl 459  
 DB 959 TATTGACGCTCCCTTTTGTAGTAATCATGCACA-----GACCGGTGAGC 915  
 QY 459 aAlaAlaLeuValGlyLeuAlaGlnAspAspProHisAlaValAspArgIleLysLeuLe 479  
 DB 914 TAGCTTTTAAGCTCTTTAGCCATGCAA-----AAAATGCC 879  
 QY 479 uTyrSerIleAlaLeuSerThrGlyGlyLeuProLeuIleTyrLeuGlyAspGluValGlu 499  
 DB 878 TGCTCAAACTTACCTTTTGATGTCGGGAAATCCATTTATTATTACTACGCGGAGAAATTGGG 819

QY 499 yThrLeuAsnAspAspTyrSerGlnAspSerAsnLysSerAspAspSerArgTyr-- 518  
Db 818 GGTGACAGGTTCT-----GGATTGACCAACAAACGCTTCCCAATCCCATGGTC 768  
QY 519 -----AlaHisArgProArgTyrAsnGluAlaLeuTyrAlaGlnArgAsnAspPr 535  
Db 767 GCGTATGAGCAGATTCCTCCCTAAGCAAAATGTCTGCTCCCGTGGACGATGAACCGC 708  
QY 535 oSerThrAlaAlaGlyGlnIleTyrGln-----G1 545  
Db 707 GACACTGACGAGGATCGGTGGTCAGCAAGAAAGCAATCCGAATTCGCTACTCAATTG 648  
QY 545 yLeuArgHisMetIleAlaValArgGlnSerAsnProArgPheAspGlyGlyArgLeu-- 564  
Db 647 GTACAAAATAATCTTCGGGTCAAGCGAAATATCTGAAATTCGCACTGCCAGGCTGAA 588  
QY 565 -----ValThrPheAsnThrAsnLysHisIleIleGlyTyrIleArgAsnAlaLe 583  
Db 587 AAATATTCGTCAGTAAGTAGCTGTCTGCTCATTAATTAATGGAAGGAT-----TT 534  
QY 583 uLeuAlaPheGlyAsnPheSerGluTyrProGlnThrValThrAlaHisThrLeuGlnAl 603  
Db 533 GACCATTGCAATAATTTTCAGT---AAAGAAAGTATGACAGTGAAACTCCCGAAGG 477  
QY 603 aMetProPheLys-----AlaHisAspLeuIleGlyLysThrValSerLeuAsnGl 621  
Db 476 AGTTGTGGCAAAATAATTAAGTACCAACTTTCTGTGTGGTGGAGTCCAGCGACATTGAC 417  
QY 621 nAsp-----LeuThrLeuGlnProTyrGlnValMetTyrLeuGlu 634  
Db 416 TGAATGGAATAATGACAAATCCAGCTTACAGTACAGTATTTTTGAAA 371

RESULT 7  
BM610365  
LOCUS  
DEFINITION  
17000687110664 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone  
19600449688657 5', mRNA sequence.

ACCESSION  
BM610365  
VERSION  
BM610365.1 GI:18908469  
KEYWORDS  
EST.  
SOURCE  
Anopheles gambiae (African malaria mosquito)  
ORGANISM  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicidae;  
Anopheles.  
REFERENCE  
1 (bases 1 to 673)  
Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab  
R., Collins,F.H., Venter,J.C. and Hoffman,S.L.  
Celera Anopheles gambiae EST project  
Unpublished  
Contact: Holt R.A.  
Celera Genomics  
45 W. Gude Dr., Rockville, MD 20850, USA  
Tel: 2404533151  
Fax: 2404534580  
Email: HoltRA@celera.com  
Plate: NU01004AY2 row: p column: 19  
Seq primer: M13 Reverse.

FEATURES  
source  
1..673  
/organism="Anopheles gambiae"  
/mol\_type="mRNA"  
/strain="RSP-ST (Reduced susc. to Permethrin - std.  
chromosome)"  
/db\_xref="taxon:7165"  
/clone="19600449688657"  
/dev\_stage="Adult"  
/lab\_host="DHI0P"  
/clone\_lib="A.Gam.ad.cDNA.blood1"  
/notes="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole  
adult mosquitoes (mixed sex) frozen on liquid nitrogen 24  
hours after human blood feeding. cDNA inserts >500 bp  
cloned directionally into pSport 1. Not 1 site is 3'.

Clones available through the Malaria Research and  
Reference Reagent Resource Center (www.malaria.mr4.org)"  
BASE COUNT 154 a 174 c 197 g 148 t  
ORIGIN  
Alignment Scores:  
2.92e-21 Length: 673  
Score: 271.50 Matches: 65  
Percent Similarity: 47.87% Conservative: 36  
Best Local Similarity: 30.81% Mismatches: 69  
Query Match: 7.97% Indels: 41  
DB: 12 Gaps: 7

US-09-843-007A-2 (1-636) x BM610365 (1-673)  
QY 118 GlyAspLeuLysGlyLeuLysAspLysIleProTyrPhe---GlnGluLeuGlyLeuThr 136  
Db 38 GGAGATTTCGCGCGCAATAATGAAAAAGTTCCGTACCTGCGACGAGCTCGCATCGAT 97  
QY 137 TyrLeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAla 156  
Db 98 GCATATGGCTGTGCGCGCATCTTCAAGTCACCG-----ATGGCCGACTTTGGCTAGCAT 151  
QY 157 ValSerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluVal 176  
Db 152 ATTCGCGATTTCGCGCATTCATTCGGAGTTTGGACAAATTCGTGACCTGGAGCGCTG 211  
QY 177 IleAlaLeuHisGluAlaGlyIleSerAlaValAlaAspPheIlePheAsnHisThr 196  
Db 212 CGGACCGCTTCTGAATGACAGAGGCTCAAGCTAATCTCTGACCTTTGTTCCGAACACAGC 271  
QY 197 SerAsnGluHisGluTyrAlaGlnArgCysAlaAlaGlyAspProLeuPheAsnAspPhe 216  
Db 272 AGCATTGAAATGAGTGGTGTCTTAAAGTCCGTCAGAGGATCCCACTACGACGATTAC 331  
QY 217 TyrTyrIlePheProAspArg-----ArgMetProAspGlnTyrAsp 230  
Db 332 TATGTGTGCATCCCGGTAAAGACGCTGCTAAACGGAGCGCGCTGCCT----- 379  
QY 231 ArgThrLeuArgGluIlePheProAspGlnHisProGlyGlyPhe---SerGlnLeuGlu 249  
Db 380 -----CCTTCGAACTGGGTGAGTGTCTTCGTT 406  
QY 250 AspGlyArgTyrValTyrThr-----ThrPheAsnSer 260  
Db 407 GGTTCAGCTCGGAGTGGACGACGTGCGAAGGAGTACTATCTCCACCAGTTCCTAGTT 466  
QY 261 PheGlnTyrAspLeuAsnTyrSerAsnProTyrValPheArgAlaMetAlaGlyGluMet 280  
Db 467 AAGCAGCCCGATCTCACTATCGCAATCCCGCGCTAGTACAAAGAAATGAAGACGTGATG 526  
QY 281 LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTyr 300  
Db 527 AGTTCTGCTTGGCAAGGCGTCAATGAGTTCGATCGATCGATCGCTGCGCTATCTTTT 586  
QY 301 LysGlnMetGlyThrSerCysGluAsnLeuPro 311  
Db 587 -----GAATCGCTGCC 598

RESULT 8  
BM646129  
LOCUS  
DEFINITION  
17000687319533 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone  
19600449637968 5', mRNA sequence.  
ACCESSION  
BM646129  
VERSION  
BM646129.1 GI:18945640  
KEYWORDS  
EST.  
SOURCE  
Anopheles gambiae (African malaria mosquito)  
ORGANISM  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
Anopheles.  
REFERENCE  
1 (bases 1 to 712)

AUTHORS Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L., Charlab  
 TITLE R., Collins, P.H., Venter, J.C. and Hoffman, S.L.  
 JOURNAL Celeris Anopheles gambiae EST project  
 COMMENT Unpublished  
 CONTACT: Holt R.A.  
 Celeris Genomics  
 45 W. Gude Dr., Rockville, MD 20850, USA  
 Tel.: 2404533351  
 Fax: 2404534580  
 Email: HoltRA@celera.com  
 Plate: NU01004AYS row: A column: 06  
 Seq primer: M13 Reverse.  
 FEATURES Location/Qualifiers  
 source  
 1..712  
 /organism="Anopheles gambiae"  
 /mol\_type="rRNA"  
 /strain="RSP-ST (Reduced susc. to Permethrin - std.  
 chromosome)"  
 /db\_xref="taxon:7165"  
 /clone="19600449637968"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B"  
 /clone\_lib="A.Gam.ad.cdna1"  
 /note="Vector: pSPori1; Site 1: SalI; Site 2: NotI; Whole  
 adult mosquitoes (mixed sex) frozen on liquid nitrogen.  
 cDNA inserts >500 bp cloned directionally into pSPori 1.  
 Not 1 site is 3'. Clones available through the Malaria  
 Research and Reference Reagent Resource Center  
 (www.malaria.war4.org)."  
 BASE COUNT 156 a 181 c 201 g 174 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 3,19e-21 Length: 712  
 Score: 271.50 Matches: 65  
 Percent Similarity: 47.87% Conservative: 36  
 Best Local Similarity: 30.81% Mismatches: 69  
 Query Match: 7.97% Indels: 41  
 DB: 12 Gaps: 7  
 US-09-843-007A-2 (1-636) x BM646129 (1-712)  
 QY 118 GlyAspLeuLysGlyLeuLysAspLysIleProTyrPhe---GlnGluLeuGlyLeuThr 136  
 DB 137 GGAGATTTTCGGCGGATTAATGGAAGTTTCGTACCTCGCAGCGAGCTCGGCATCGAT 196  
 QY 137 TyrLeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyTyrAla 156  
 DB 137 GCATATGCTGCTGCCGATCTTCAAGTCACG-----ATGGCGGACTTTGGCTACGAT 250  
 QY 157 ValSerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluVal 176  
 DB 251 ATTGCGGATTTTCGGCGGATTAATGGAAGTTTCGTACCTCGCAGCGAGCTCGGCATCGAT 310  
 QY 177 IleAlaAlaLeuHisGluAlaGlyIleSerAlaValValAspPheIlePheAsnHisThr 196  
 DB 311 CGGACCGTTGTAATGCGAGAGGGCTCAAGTAATCTCGGACTTTGTTCCGACACACAGC 370  
 QY 197 SerAsnGluHisGluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPhe 216  
 DB 371 AGCGATGAAGTGAAGTGGTTCTAAAGTCGCTCCAGAGGATCCACCTACAGCGATTAC 430  
 QY 217 TyrTyrIlePheProAspArg-----ArgMetProAspGlnTyrAsp 230  
 DB 431 TATGTGTGCGACCCCGGTAAGACGCTGTGCTAACCGGACGCGGTCCT----- 478  
 QY 231 ArgThrLeuArgGluIlePheProAspGlnHisProGlyGlyPhe---SerGlnLeuGlu 249  
 DB 479 -----CCTTCGAACTGGGTGAGTCTTCCGT 505  
 QY 250 AspGlyArgTyrValTyrThr-----ThrPheAsnSer 260  
 DB 506 GGTTCAGCGCTGGGAGTGAAGACGTCGGGGAAGGAGTACTATCTCCACGATTCCTGGTT 565

QY 261 PheGlnTrpAspLeuAsnTyrSerAsnProTyrPheValPheArgAlaMetAlaGlyGluMet 280  
 DB 566 AAGCAGCCCGATCTGAACACTATCCGCAATCCCGCTAGTACAGAAATCAAGATCTGA 625  
 QY 281 LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIlePhe 300  
 DB 626 ACGTTCGTGGCTTGGCAAGGGCTTCATGGATTCGCAATCGATCGCGTGTCTTCTT 685  
 QY 301 LysGlnMetGlyThrSerCysGluAsnLeuPro 311  
 DB 686 -----GAATCGTGCCT 697  
 RESULT 9  
 BZ578463 785 bp DNA linear GSS 17-DEC-2002  
 LOCUS msh2.5885.x1 msh Pseudomonas aeruginosa genomic clone msh2.5885,  
 DEFINITION genomic survey sequence.  
 ACCSSION BZ578463  
 VERSION BZ578463.1 GI:27213524  
 KEYWORDS GSS.  
 SOURCE Pseudomonas aeruginosa  
 ORGANISM Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.  
 REFERENCE 1 (bases 1 to 785)  
 AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,  
 Burns, J.L., Kaul, R. and Olsen, M.V.  
 TITLE Whole-Genome-Sequence variation among multiple isolates of  
 Pseudomonas aeruginosa library  
 JOURNAL J. Bacteriol., (2002) In press  
 COMMENT Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062218554  
 Fax: 2066857244  
 Email: craymond@u.washington.edu  
 Class: shotgun.  
 FEATURES Location/Qualifiers  
 source  
 1..785  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="MSH"  
 /db\_xref="taxon:287"  
 /clone="msh2.5885"  
 /clone\_1lb="msh"  
 /notes="Environmental isolate. Whole genomic shotgun  
 library."  
 BASE COUNT 154 a 274 c 224 g 131 t 2 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 2,2e-20 Length: 785  
 Score: 265.00 Matches: 74  
 Percent Similarity: 45.78% Conservative: 40  
 Best Local Similarity: 29.72% Mismatches: 96  
 Query Match: 7.78% Indels: 40  
 DB: 29 Gaps: 10  
 US-09-843-007A-2 (1-636) x BZ578463 (1-785)  
 QY 174 ArgGluValIleAlaAlaLeuHisGluAlaGlyIleSerAlaValValAspPheIlePhe 193  
 DB 90 CGCGGTTCATCCCGAGGCGCCATCGACGGTCTCGGGGGATTACCGAGCTGTGATC 149  
 QY 194 AsnHisThrSerAsnGluHisGluTrp-----AlaGlnArgCysAlaAlaGlyAsp 210  
 DB 150 AACCACTTCGACGACGATCCCTGGTTCATCCGCGCCCGCAGGAGGATCG 209  
 QY 211 ProLeuPheAspAsnPheTyrIlePheProAspArgMetProAspGlnTyrAsp 230  
 DB 210 CGCGCCCGCGAC-----TACTAGTCTGTGCGAC-----ACCGACGAGAAATACCG 257

```

QY 231 ArgThrLeuArgGluLeuPheProAspGlnHisProGlyGlyPheSerGlnLeuGluAsp 250
DB 258 GGTACG---CGGATCATCTTCATCGAC-----ACCGAGCAG 290
QY 251 GlyArgTrpValTrpThr-----ThrPheAsnSerPhe 261
DB 291 TCCAACCTGGACCTGGACCGGTGGCCACACACTACTACGACCGCTTCATTCCAC 350
QY 262 GlnTrpAspLeuAsnTrpSerAsnProTrpValPheArgAlaMetAlaGlyGluMetLeu 281
DB 351 CAGCGCGACCTGAACCTCGACCAACCGGAGTCTCGCGAGGTCTCGGGGTGATGGC 410
QY 282 PheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaValAlaPheIleTrpLys 301
DB 411 TACTGGCTGGACATGGCGTCGACCGCTCGCGCTGACCGGATTCCTACCTGATCGAA 470
QY 302 GlnMetGlyThrSerCysGluAsnLeuProGlnAlaHisAlaLeuIleArgAlaPheAsn 321
DB 471 CGCGACGGACACGAGCAGGAGAACCTGCGGAGACCCACGAGTGTCTCAGCGCATCGC 530
QY 322 AlaValMetArgIleAlaAlaProAlaValPhePheLysSerGluAlaIleValHisPro 341
DB 531 GCGGAGCTGGACGGCGCACTATCCGACCGCATCTGTCGCGGAGGCCAACCGATGGCG 590
QY 342 AspGlnValValGlnTrpIleGly-----GlnAspGluCysGlnIle 355
DB 591 GAGACACCGCGCGGACTCTCGCTGCGAGGATGGCGGTAGCGGACGAGATCCCATG 650
QY 356 GlyTyrAsn---ProLeuGlnMetAlaLeuLeuTrp-----AsnThrLeuAlaThr 371
DB 651 GCCTTTCACCTCCGCTGATCGCGCATGTACATGGCCATCCGCCAGGAGATCGCTATC 710
QY 372 ArgValAlaAsnLeuHisGlnAlaLeuThrTyrArgHisAsnLeuProGluHisThr 391
DB 711 CGATC-ACCGACATCTCGCGCCAGACCTC-----GGCATTCGCTCCAAATGC 757
QY 392 AlaTrpValAsnTrpValArgSerHis 400
DB 758 CAATGGCGCATCTTCTCGCCCAACAC 784

RESULT 10
BM619567 587 bp mRNA linear EST 25-FEB-2002
LOCUS 17000687375124 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
DEFINITION 19600449632068 5', mRNA sequence.
ACCESSION BM619567
VERSION BM619567.1 GI:189517985
KEYWORDS EST.
SOURCE Anopheles gambiae (African malaria mosquito)
ORGANISM
REFERENCE 1 (bases 1 to 587)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
R., Collins,F.H., Venter,J.C. and Hoffman,S.I.
TITLE Ceiera Anopheles gambiae EST project
JOURNAL Unpublished
COMMENT Contact: Holt R.A.
Ceiera Genomics
45 W. Gude Dr.
Rockville, MD 20850, USA
Tel: 2404531151
Fax: 2404534580
Email: HoltRA@ceiera.com
Plate: NU01004N8D row: N column: 14
Seq primer: M13 Reverse.
Location/Qualifiers
FEATURES
source
1..587
/organism="Anopheles gambiae"
/mol_type="mRNA"
/strain="ESP-ST (Reduced susc. to Permethrin - std.
chromosome)"

```

```

/db_xref="taxon:7165"
/clone="19600449622928"
/dev_stage="Adult"
/lab_host="DH10B"
/clone_lib="A.Gam.ad.cDNA1"
/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."
BASE COUNT 135 a 151 c 168 g 133 t
ORIGIN

```

```

Alignment Scores: 2,41e-20 Length: 587
Pred. No.: 263.00 Matches: 62
Score: 47.96% Conservative: 32
Percent Similarity: 31.63% Mismatches: 68
Best Local Similarity: 7.72% Indels: 34
Query Match: 12 Gaps: 6
DB:

```

US-09-843-007a-2 (1-636) x BM619567 (1-587)

```

QY 118 GlyAspLeuLysGlyLeuLysAspLysIleProTyrPhe---GlnGluLeuGlyLeuThr 136
DB 44 GGAGATTTCCGCGGCAATAATGGAAAAGTTCCGTACCTGCGACGGAGCTCGGCATCGAT 103
QY 137 TyrLeuHisLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAla 156
DB 104 GCATATGCTGTCGCGCATCTTCAAGTCACCG-----ATGCCGACTTTGGTACGAT 157
QY 157 ValSerSerTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluVal 176
DB 158 ATTCGCGATTTCCGCGACATTCATTCGGAGTTTGGACAAATTCGTGACCTGGAAGCGCTG 217
QY 177 IleAlaLeuHisGluAlaGlyIleSerAlaValAlaAspPheIlePheAsnHisThr 196
DB 218 GGCACGCGCTTGAATGCAGAGCGGCTCAAGCTAACTCTGGACTTTGTCGAAACACACAGC 277
QY 197 SerAsnGluHisGluTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPhe 216
DB 278 AGCGATGAAGTAGTGTTTCTTAAGTCCGTCAGAGGATCCACCTACAGCGATTAC 337
QY 217 TyrTyrIlePheProAspArg-----ArgMetProAspGlyTyrAsp 230
DB 338 TATGTGTGGCATCCGCGTAAGACGCTGGCTAACGGGACGCGCGTGCCT-----385
QY 231 ArgThrLeuArgGluIlePheProAspGlnHisProGlyGlyPhe---SerGlnLeuGlu 249
DB 386 -----CCTTCGACTGGTGGTGTCTCCCGT 412
QY 250 AspGlyArgTrpValTrpThr-----ThrPheAsnSer 260
DB 413 GGTTCAGCTCGGAGTGGAAACGACGTCGGAAGAGTACTATCTCCACGATTCCTAGTT 472
QY 261 PheGlnTrpAspLeuAsnTrpSerAsnProTrpValPheArgAlaMetAlaGlyGluMet 280
DB 473 AAGCAGCCCGATCTGAACATATCGCATCCCGCTAGTACAGAAATGAAGACGCTGATG 532
QY 281 LeuPheLeuAlaAsnLeuGlyValAspIleLeuArgMetAspAlaVal 296
DB 533 ACCTTCGCTGGCGAAGGCGTCCATGATTCGATCGATCGATCGCGTG 580

```

RESULT 11

```

BM652118 711 bp mRNA linear EST 26-FEB-2002
LOCUS 17000687375124 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
DEFINITION 19600449632068 5', mRNA sequence.
ACCESSION BM652118
VERSION BM652118.1 GI:18951629
KEYWORDS EST.
SOURCE Anopheles gambiae (African malaria mosquito)

```

Dd	516	-----CCTTCGAACTGGGTGAGTCTCCCGT 542
Qy	250	AspGlyArgTrpValTrpThr -----ThrpheAnSer 260
Dd	543	GTTTCACGCTGGGAGTGGAACGACGCGCGAAGAGTAGTAATACTATCTCCACCAGTTCTCTAGTT 602
Qy	261	PheGlnTrpAspLeuAnStryrSerAnProTrpValPheArgAlaMetAlaGlyGluMet 280
Dd	603	AAGCACGCCGATCTGAACATATCGCAATCCCGCTAGTACAGAAGATTGAAGACGGTGATG 662
Qy	281	LeuPheLeuAlaAnSleuGlyValAspIleLeuArgMetAspAlaVal 296
Dd	663	ACGTTCTGCTTGGCAAAGGGGTCCTCATCGCATCGATCGATCGCGTG 710
RESULT 12		
BM628529		
LOCUS	704 bp	mRNA linear EST 26-FEB-2002
DEFINITION	170006874968686 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone	
	196004499832435 5', mRNA sequence.	
ACCESSION	BM628529	
VERSION	BM628529.1	GI:18928040
KEYWORDS	EST.	
SOURCE	Anopheles gambiae (African malaria mosquito)	
ORGANISM	Anopheles gambiae	
	Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;	
	Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;	
	Anopheles.	
REFERENCE	1 (bases 1 to 704)	
AUTHORS	Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab,	
	R., Collins,F.H., Venter,J.C. and Hoffman,S.L.	
TITLE	Celera Anopheles gambiae EST project	
JOURNAL	Unpublished	
COMMENT	Contact: Holt R.A. Celera Genomics 45 W. Gude Dr., Rockville, MD 20850, USA Tel: 2404533151 Fax: 2404534580 Email: HoltRA@celera.com Plate: NU010049VT row: J column: 17 Seq primer: M13 Reverse.	
FEATURES	Location/Qualifiers	
source	1..704	
	/organism="Anopheles gambiae"	
	/mol_type="mRNA"	
	/strain="RSP-ST (Reduced susc. to Permethrin - std.	
	chromosome)"	
	/db_xref="taxon:7165"	
	/clone="19600449632435"	
	/dev_stage="Adult"	
	/lab_host="DH10b"	
	/clone_lib="A.Gam.ad.cDNA1"	
	/note="Vector: pSport; Site 1: SalI; Site 2: NotI; Whole	
	adult mosquitoes (mixed sex) frozen on liquid nitrogen.	
	cDNA inserts >500 bp cloned directionally into pSport 1.	
	Not 1 site is 3' Clones available through the Malaria	
	Research and Reference Reagent Resource Center	
	(www.malaria.rri.org)."	
BASE COUNT	160 a 162 c 213 g 169 t	
ORIGIN		
Alignment Scores:		
Pred. No.:	3,67e-20	Length: 704
Score:	262.50	Matches: 68
Percent Similarity:	42.38%	Conservative: 46
Best Local Similarity:	25.28%	Mismatches: 77
Query Match:	7.70%	Indels: 78
DB:	12	Gaps: 10
US-09-843-007A-2 {1-636} x BM628529 (1-704)		
Qy	59	TyrGlyAsnScnclulalaLeuLeuPrometLeuGluMerLeuLeuAlaGlnAfaTrp--- 77
Dd	19	CATGGGAATCTCCAAA-----CCATCTGCAGTTGTGTGTAGCTGTGTGGGC 69

78 -----GlnSerTyrSerGlnArgAsnSerSerLeuLysAspIle 90  
 70 ACTCTTTAGTGGCATGTTAGATCTTAC----- 99  
 91 AspIleAlaArgGluAsnAsnProAspTrpIleLeuSerAsnLysGlnValGlyGlyVal 110  
 100 -----CGTACCAACGTCCTCG-----CGATTGTGGGAAAGGC 132  
 111 Cys-TyrValAspLeuPhe-----AlaGlyAs 119  
 133 TGCCTTCTATCAATATATCCCGATCGTTCATGACAGTGTGGGATGTTGGTGGCGCA 192  
 119 pLeuLysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHi 139  
 193 TTTTAAAGGAATGCTTCAAGCTCGTACCTGAAGTCGATCGAGTGAAGCGCTTCTG 252  
 139 sLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSe 159  
 253 GATGTCACCGATCTCAAAATCACCG-----ATGTCGATTTTGGCTATGATATTCTGA 306  
 159 rTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAl 179  
 307 CTTTCGGGACATCCAGAGATTGGACGATGCTGATTTGACCGTTCGGTGAGCA 366  
 179 aLeuHisGluAlaGlyIleSerAlaValAspPheIlePheAsnHisThrSerAsnGl 199  
 367 GCGCACCGGCTCGGCTCGAGTGTATGATGATCTTGTCCGAAACCATTCGAGCAATCT 426  
 199 uHisGlnTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTril 219  
 427 GCACGAGTGTTCGTAAGTCGGAACGACGAGCGCGGCTATGAGGAC---TATTACGT 483  
 219 ePheProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAs 239  
 484 GTGGCATGATCGAAGTT----- 502  
 239 pGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArgTrpValTrpThrPheAs 259  
 503 ----AATCGCGCGCGGTGAGGACTTACCACCAAAATTCGATTCAGGCTTCGCTG 558  
 259 nSer---PheGlnTrp----- 263  
 559 CAGTGCATCGCATGGAGTGATCAGCGGCAACAGTACTACCTACACCGATTCACCGTCA 618  
 264 -----AspLeuAsnTyrSerAsnProTrpValPheArgAlaMetAlaGlyGluMetle 281  
 619 GCAACCGGACTTAAACTACGAAATCCGACCGTTCGTTACGAGATGAAGATGTCTGCT 678  
 281 uPheLeuAlaAsnLeuGlyValAsp 289  
 679 GTTCTGGTGGCGAAGGCGTGGAC 703

BM650445 706 bp mRNA linear EST 26-FEB-2002  
 LOCUS 17000687371299 A.Gam.ad.cdNAL Anopheles gambiae cdNA clone  
 DEFINITION 19600449650921 5', mRNA sequence.  
 ACCESSION BM650445  
 VERSION BM650445.1 GI:18949956  
 KEYWORDS EST.  
 SOURCE Anopheles gambiae (African malaria mosquito)  
 ORGANISM Anopheles gambiae  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
 Anopheles.  
 1 (bases 1 to 706)  
 AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab  
 ,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.  
 TITLE Celera Anopheles gambiae EST project  
 JOURNAL Unpublished  
 COMMENT Contact: Holt R.A.  
 Celera Genomics

45 W. Gude Dr., Rockville, MD 20850, USA  
 Tel: 2404533151  
 Fax: 24045334580  
 Email: Holt@celera.com  
 Plate: NU010049W7 row: L column: 23  
 Seq primer: M13 Reverse.  
 Location/Qualifiers

## FEATURES

source 1..706  
 /organism="Anopheles gambiae"  
 /mol\_type="mRNA"  
 /strain="RSP-ST (Reduced susc. to Permethrin - std. chromosome)"  
 /db\_xref="taxon:7165"  
 /clone="19600449650921"  
 /dev\_stage="Adult"  
 /lab\_host="DH10b"  
 /clone\_lib="A.Gam.ad.cdNAL"  
 /note="Vector: pSport1; Site\_1: SalI; Site\_2: NotI; Whole adult mosquitoes (mixed sex) frozen on liquid nitrogen. cDNA inserts >500 bp cloned directionally into pSport 1. Not 1 site is 3'. Clones available through the Malaria Research and Reference Reagent Resource Center (www.malaria.mr4.org)."

BASE COUNT 160 a 162 c 215 g 169 t  
 ORIGIN

## Alignment Scores:

Score: 3.69e-20 Length: 706  
 Pzed. No.: 262.50 Matches: 68  
 Percent Similarity: 42.38% Conservative: 46  
 Best Local Similarity: 25.28% Mismatches: 77  
 Query Match: 7.70% Indels: 78  
 DB: 12 Gaps: 10

US-09-843-007A-2 (1-636) x BM650445 (1-706)

QY 59 TyrGlyAsnAsnGluAlaLeuLeuProMetLeuGluMetLeuLeuAlaGlnAlaTrp--- 77  
 Db 19 CATGGGACTCCAAA-----CCCATCTGATGTGTGTGTAGCCCTTGTCTGGGCG 69  
 QY 78 -----GlnSerTyrSerGlnArgAsnSerSerLeuLysAspIle 90  
 Db 70 ACTCTTTTAGTGGCATGTTTAGATCTTAC----- 99  
 QY 91 AspIleAlaArgGluAsnAsnProAspTrpIleLeuSerAsnLysGlnValGlyGlyVal 110  
 Db 100 -----CGTACCAACGTCCTCG-----CGATTGTGGGAAAGGC 132  
 QY 111 Cys-TyrValAspLeuPhe-----AlaGlyAs 119  
 Db 133 TGCCTTCTATCAATATATCCCGATCGTTCATGACAGTGTGGGATGTTGGTGGCGCA 192  
 QY 119 pLeuLysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHi 139  
 Db 193 TTTTAAAGGAATGCTTCAAGCTCGTACCTGAAGTCGATCGAGTGAAGCGCTTCTG 252  
 QY 139 sLeuMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSe 159  
 Db 253 GATGTCACCGATCTCAAAATCACCG-----ATGTCGATTTTGGCTATGATATTCTGA 306  
 QY 159 rTyrArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaAl 179  
 Db 307 CTTTCGGGACATCCAGAGATTGGACGATGCTGATTTGACCGTTCGGTGAGCA 366  
 QY 179 aLeuHisGluAlaGlyIleSerAlaValAspPheIlePheAsnHisThrSerAsnGl 199  
 Db 367 GCGCACCGGCTCGGCTCGAGTGTATGATGATCTTGTCCGAAACCATTCGAGCAATCT 426  
 QY 199 uHisGlnTrpAlaGlnArgCysAlaAlaGlyAspProLeuPheAspAsnPheTyrTril 219  
 Db 427 GCACGAGTGTTCGTAAGTCGGAACGACGAGCGCGGCTATGAGGAC---TATTACGT 483  
 QY 219 ePheProAspArgArgMetProAspGlnTyrAspArgThrLeuArgGluIlePheProAs 239

```

Db      484 GTGCATGATGGGAAGTT----- 502
Qy      239 pGlnHisProGlyGlyPheSerGlnLeuGluAspGlyArgTrpValTrpThrPheAs 259
Db      503 -----AATCGCGCGGGCGGTAGGAACCTTACCACCAAAACATTTGGATTCAAGCCTTCGCGTG 558
Qy      259 nSer---PheGlnTrp----- 263
Db      559 CAGTCATGTCAGTGGAGTATCAGCGGCAACAGTACTACTACACAGTTTACCGTCGA 618
Qy      264 -----AspLeuAsnTyrSerAsnProTrpValPheArgAlaMetalGlyGluMetLe 281
Db      619 GCACCGGACTTAAATACCGAATCCGACGCGTTCAGGAGATGAAGATGTCGTGCT 678
Qy      281 uPheLeuAlaAsnLeuGlyValAsp 289
Db      679 GTTCTGGCTGGCAAGGGGTGGAC 703

RESULT 14
CNS08PQT
LOCUS
DEFINITION
Single read from an extremity of a full-length cDNA clone made from
Anopheles gambiae total adult females. 5-PRIME end of clone
FK0AA34DD06 of strain 6-9 of Anopheles gambiae (African malaria
mosquito).
ACCESSION
BX022785
VERSION
BX022785.1 GI:27572005
KEYWORDS
HTC.
SOURCE
Anopheles gambiae (African malaria mosquito)
ORGANISM
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.
REFERENCE
1 (bases 1 to 912)
AUTHORS
Direct Submission
JOURNAL
Submitted (06-JAN-2003) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
FEATURES
Location/Qualifiers
1..912
/organism="Anopheles gambiae"
/mol_type="cDNA"
/strain="6-9"
/db_xref="taxon:7165"
/clone="FK0AA34DD06"
/plasmid="pME18S-FL"
/note="end : 5-PRIME"
BASE COUNT 214 a 231 c 290 g 177 t
ORIGIN

Alignment Scores:
Pred. No.: 5,53e-20 Length: 912
Score: 262.50 Matches: 87
Percent Similarity: 43.61% Conservative: 53
Best Local Similarity: 27.10% Mismatches: 115
Query Match: 7.70% Indels: 66
DB: 11 Gaps: 12

US-09-843-007a-2 (1-636) x CNS08PQT (1-912)
Qy      76 AlaTrpGlnSer-----TyrSerGlnArgAsnSerSerLeuLeuAspIleAspIle 92
Db      79 GCGTGGGATCGGGCGGTGTTTACCAGATCTATCCGCGATCTTCAGGAC----- 129
Qy      93 AlaArgGluAsnAsnProAspTrpIleLeuSerAsnLysGlnValGlyValCysTyr 112
Db      130 -----AGCAATGGGACGGTGTG----- 147
Qy      113 ValAspLeuPheAlaGlyAspLeuLysGlyLeuLysAspLysIleProTyrPheGlnGlu 132
Db      148 -----GGCATCTGACCGGCGCATTCAGGAGAGCTGGAGACATTTGGCCGAT 192

```

```

Qy      133 LeuGlyLeuThrTyrLeuHisLeuMetMetProLeuPheLysCysProGluGlyLysSerAsp 152
Db      193 CTGTGTGTGACTGGGTGTGGCTGAGCCGGTGTGTTAAATCACC-----ATGGCGGAC 246
Qy      153 GlyGlyTyrAlaValSerSerTyrArgAspValAsnProAlaLeuGlyTrpIleGlyAsp 172
Db      247 TTGGGTAGCATATTTCGGACTTCGGCATTCGGCATTCGGATTCGAATCTTTGGCACTATGGCCGAT 306
Qy      173 LeuArgGluValIleAlaAlaLeuHisGluAlaGlyIleSerAlaValValAspPheIle 192
Db      307 TTGGACCGTATGTGTCAGAAAGGCAAGACGCTCGGTATTAAAGTGATCTTGGACTTTGTG 366
Qy      193 PheAsnHisSerAsnGluHisGlnTrpAlaArgCysAlaAlaGlyAspProLeu 212
Db      367 CCGAACCATACAGGCGAGCGACGACGACGAGTGTGTTGTAAGTCGCTGAACACGAGGAGAC 426
Qy      213 PheAspAsnPheTyrTyrIlePhe-----ProAspArgArgMetProAspGlnTyr 229
Db      427 TACCGGAT---TACTATGTGTGGCGCAACGGTGTGAACGGTGAACGCCCAACAAATTGG 483
Qy      230 AspArgThrLeuArgGluIlePheProAsp-----GlnHisProGlyGlyPheSer 246
Db      484 -----CAATCCGTGTTCCATACGCCCGCTGGACCGACGCTGCCCGCAACAAACG 531
Qy      247 GlnLeuGluAspGlyArgTrpValTrpThrPheAsnSerPheGlnTrpAspLeuAsn 266
Db      532 CAG-----TACTATCTGCACCATGTTTCGACAAAGACGACGCGGATCTGAAC 576
Qy      267 TyrSerAsnProTrpValPheArgAlaMetAlaGlyGluMetLeuPheLeuAlaAsnLeu 286
Db      577 TACCCCAACCCGAAATGAAGCAGGAGATGGCGGACATGGTCCGCTTTCTGGCTGGACAAAG 636
Qy      287 GlyValAspIleLeuArgMetAspAlaValAlaPheIleTrpLysGlnMetGlyThrSer 306
Db      637 GGCATCGAAGGGTTCGCGATCGATCGATCAACACGAGTGTAGCGGATCCCCAGTTCGT 696
Qy      307 CysGluAsnLeu-----ProGlnAlaHis 314
Db      697 GACGAGGAGCTGATCGATCCGAAGGTGAGCTGATCTGGGAAACCTCGCACCAAGTACAC 756
Qy      315 Ala---LeuIleArgAlaPheAsnAlaValMetArgIleAlaAlaProAlaValPhePhe 333
Db      757 GCAGAACCTGCCGAGTCTAGATCTCATCTACACTGGCG----- 798
Qy      334 LysSerGluAlaIleValHisProAspGln-ValValGlnTyrIleGlyGlnAspGluCys 353
Db      799 -----CGATGTGTTCGACCACTACAGCGCGCAGCAATGTGACGCGCTGATGAT 849
Qy      353 sGlnIleGlyTyrAsnProLeuGlnMetAlaLeuLeuTrpAsnThrLeuAlaThrArgG1 373
Db      850 GACGGAACGTACGCAACCTCGAGCAGACGATCTCTGTGACGCAATGCGACCGCA 909
Qy      373 u 373
Db      910 G 910

RESULT 15
BM579790
LOCUS
DEFINITION
17000687241299 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone
19600449716264 5', mRNA sequence.
ACCESSION
BM579790
VERSION
BM579790.1 GI:18868257
KEYWORDS
EST.
SOURCE
Anopheles gambiae (African malaria mosquito)
ORGANISM
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.
REFERENCE
1 (bases 1 to 694)
AUTHORS
Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
R., Collins,F.H., Venter,J.C. and Hoffman,S.L.

```

Db 421 CGAGTGGTTCGTAAGTCGGACAGCGAGCGCGGCTATGAGGAC ---TATTACGTGTG 477  
Qy 220 eProAspArgMetProAspGlnTyrAspArgThrLeuArgGluLeuPheProAspG1 240  
Db 478 GCATGATGGGAAGGTT----- 493  
Qy 240 nHisProGlyGlyPheSerGlnLeuGluAspGlyArgTyrValTyrThrPheAsnSe 260  
Db 494 -AATCCGGCGCGGTAGGAACCTTACCACCAACAATGGATTCAAGCCTTCCGTGGCAG 552  
Qy 260 r---PheGlnTyr----- 263  
Db 553 TGCATGGCAGTGGAGTGCATCAGCGGCAACAGTACTACTACACCAAGTTCACCGTCGAGCA 612  
Qy 264 ---AspLeuAsnTyrSerAsnProTyrValPheArgAlaMetAlaGlyGluMetLeuPh 282  
Db 613 ACCGACTTAACTTACCGAATCCGAGTCCGCTTCAGAGATGAAGATGTCTGTCTGT 672  
Qy 282 eLeuAlaAsnLeuGlyValAsp 289  
Db 673 CTGGCTGGGCAAGGCGGTGGAC 694

Search completed: November 8, 2003, 22:40:48  
Job time : 2895 secs

TITLE Celera Anopheles gambiae EST project  
JOURNAL Unpublished  
COMMENT Contact: Holt R.A.  
Celera Genomics  
45 W. Gude Dr., Rockville, MD 20850, USA  
Tel: 2404533151  
Fax: 2404534580  
Email: HoltRA@celera.com  
Plate: NU01004HLX row: O column: 02  
Seq primer: M13 Reverse.  
Location/Qualifiers  
source 1. .694  
/organism="Anopheles gambiae"  
/mol\_type="mRNA"  
/strain="RSP-ST (Reduced susc. to Permethrin - std.  
chromosome)"  
/db\_xref="taxon:7165"  
/clone="19600449716264"  
/dev\_stage="Adult"  
/lab\_host="DH10b"  
/clone\_lib="A.Gam.ad.cDNA.bloodi"  
/notes="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole  
adult mosquitoes (mixed sex) frozen on liquid nitrogen 24  
hours after human blood feeding. cDNA inserts >500 bp  
cloned directionally into pSport 1. Not 1 site is 3'.  
Clones available through the Malaria Research and  
Reference Reagent Resource Center (www.malaria.mr4.org)"  
BASE COUNT 157 a 160 c 210 g 167 t  
ORIGIN

Alignment Scores:  
Pred. No.: 6.2e-20 Length: 694  
Score: 260.50 Matches: 68  
Percent Similarity: 42.16% Conservative: 45  
Best Local Similarity: 25.37% Mismatches: 77  
Query Match: 7.64% Indels: 78  
DB: 12 Gaps: 10

US-09-843-007A-2 (1-636) x BM579790 (1-694)

Qy 60 GlyAsnAsnGluAlaLeuLeuProMetLeuGluMetLeuLeuAlaGlnAlaTyr----- 77  
Db 13 GGGAACTCCAAA-----CCCATCTGGATGTGTGTGTAGCCCTTCTGGGCACT 63  
Qy 78 -----GlnSerTyrSerGlnArgAsnSerSerLeuLysAspIleAsp 91  
Db 64 CTCTTTAGTGGCANGTTTAGATCTTAC----- 90  
Qy 92 IleAlaArgGluAsnAsnProAspTyrIleLeuSerAsnLysGlnValGlyGlyValCys 111  
Db 91 -----CGTACCACGTCGCG-----CGATTGTGGGAAAAGGCTGC 126  
Qy 112 -TyrValAspLeuPhe-----AlaGlyAspLe 120  
Db 127 CTCTATCAAAATCTATCCCGATCGTTTCATGGACAGTGGGATGGTGTGGCGATCT 186  
Qy 120 uLysGlyLeuLysAspLysIleProTyrPheGlnGluLeuGlyLeuThrTyrLeuHisLe 140  
Db 187 TAACGGNATTCCTCGNAGCTGCGTACCTGAAGTCGATCGATGGAGCGGTTCTGGAT 246  
Qy 140 uMetProLeuPheLysCysProGluGlyLysSerAspGlyGlyTyrAlaValSerSery 160  
Db 247 GTCACCGCATCTACAAATCAGCG-----ATGCTGATTTTGGCTATGATATTCTGACT 300  
Qy 160 rArgAspValAsnProAlaLeuGlyThrIleGlyAspLeuArgGluValIleAlaLe 180  
Db 301 TCGGACATTCACAGAGAGTTTGAACGATGGCTGACTTTGACGGGTTGGTGAGCAGGC 360  
Qy 180 uHisGluAlaGlyIleSerAlaValAspPheIlePheAsnHisThrSerAsnGluHi 200  
Db 361 GCACGGGCTCGGCTGAAGGTGATCATGAGACTTTGTCCGCAACCAATTCGAGCAATCTGCA 420  
Qy 200 sGluTyrPalagLncArgCysAlaAlaGlyAspProLeuPheAsnPheTyrTyrIlePh 220